HEURISTIC INTELLIGENT METHOD FOR MACHINE MAINTENANCE AND PROCESS OPTIMIZATION

A Dissertation Proposal
Submitted to the Faculty of the
George W. Woodruff School of Mechanical Engineering

by

Yanfei Lu

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
George W. Woodruff School of Mechanical Engineering

Georgia Institute of Technology
August 2019

HEURISTIC INTELLIGENT METHOD FOR MACHINE MAINTENANCE AND PROCESS OPTIMIZATION

Approved by:

Dr. Steven Y. Liang
School of Mechanical Engineering
Georgia Institute of Technology

Dr. Thomas G. Habetler
School of Electrical and Computer Engineering
Georgia Institute of Technology

Dr. Thomas R. Kurfess
School of Mechanical Engineering
Georgia Institute of Technology

Dr. J. Kurt Jacobus
President & CEO
MedShape Inc.

Dr. Shreyes N. Melkote
School of Mechanical Engineering
Georgia Institute of Technology

Date Approved: July 25, 2019

To my parents,

Thank you for your enormous support.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. Steven Y. Liang for providing me the platform for my Ph.D. study. His encouragement and guidance have motivated me through my Ph.D. study at Georgia Tech. I will not be able to make as much accomplishment without his continuous support and guidance during my Ph.D. study. His advice and believe in my capability are necessities for me to complete my publication in various journals. His dedication to work and research activities significantly influenced me and my future career development. I really appreciated all the help and opportunities he has provided.

I also want to thank my parents for their unlimited support and encouragement during my study.

I want to thank my committee member Dr. Thomas Kurfess, Dr. J. Kurt Jacobus, Dr. Shreyes N. Melkote, Dr. Thomas G. Habetler for their advice and help of providing their valuable insights and comments. Additionally, I would like to thank Dr. Kurfess for encouraging me to attend graduate school. I would like to thank Dr. J. Kurt Jacobus for providing career advice, business insights and support of various aspects over the years.

Next, I would like to thank for my colleagues, Dr.Zhipeng Pan, Dr.Manik Rajora, Dr.Rui Xie, Dr.Yixuan Feng, Dr.Caixu Yue, Vinh Nguyen. Their help and suggestions have inspired me with more ideas for scientific publications to make this dissertation possible. I would like to thank Dr.Rui Xie for providing great innovative ideas for publications. I would like to thank Dr.Zhipeng Pan for his guidance on finite element simulations.

I want to thank my friend and mentor Jeremy W. Blair for providing extensive support. His advice and encouragement significantly helped me through my Ph.D. study. I would like to thank Jialin Zhang, Dr.Xingjian Wang and Dr.Yanxing Wang for helping me through my roughest time. I want to thank Jialin Zhang especially for her enormous support and advice. I want to thank Don Griffin, Jack Griffis, Dr.Kenneth A. Gall and Dr.Felmont Eaves III for providing me the opportunity to learn about interesting industrial applications. I would also like to thank my friend Dr.David R. Lynn, Andrew R. Castillo, Chris Longrey, He Zhu, Chang Yu, Yunpei Yang, Dr.Aoyu Chen, Chong Ye, Paul Guinee, Ilya Dolgopyat, and April Thomas.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

AIC	Akaike Information Criterion
AR	Auto Regressive
ARCH	Auto-Regressive Conditionally Heteroscedastic
ARIMA	Auto-Regressive Integrated Moving-Average
ARMA	Auto Regressive and Moving Average
BPFI	Ball Pass Frequency, Inner Race
BPF	Ball Pass Frequency
BPFO	Ball Pass Frequency, Outer Race
CBM	Condition-Based Maintenance
CCD	Charge Coupled Device
CEEMD	Complementary Ensemble Empirical Mode Decomposition
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DTWT	Dual-Tree Wavelet Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
ECM	Electro-Chemical Machining
EDM	Electrical Discharge Machining
EEMD	Ensemble Empirical Mode Decomposition
EKF	Extended Kalman Filter
EMD	Empirical Mode Decomposition
EWT	Empirical Wavelet Transform

FFT	Fast Fourier Transform
FTF	Fundamental Train Frequency
GMM	Gaussian Mixture Model
HAZ	Heat-Affected Zone
HFRT	High Frequency Resonance
HMM	Hidden Markov Model
IMF	Intrinsic Mode Function
KPIs	Key Performance Indicators
K-SVD	K Means Singular Value Decomposition
LBM	Laser Beam Machining
LMS	Least Mean Square
LP	The Lundberg and Palmgren Model
MA	Moving Average
MAE	Mean Average Error
MAPE	Mean Average Percentage Error
MAP	Maximum A Posteriori
MCA	Morphological Component Analysis
ML	Machine Learning
MSE	Mean Square Error
NN	Neural Network
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RLS	Recursive Least Square
RMS	Root Mean Square

RUL	Remaining Useful Life
RV	Realized Volatility
SE	Square Envelope
SES	Square Envelope Spectrum
SK	Spectral Kurtosis
SNR	Signal-to-Noise Ratio
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
TQWT	Tunable Q-factor Wavelet Transform
TSA	Time Synchronous Averaging
UKF	Unscented Kalman Filter
VFF	Variable Forgetting Factor
WNN	Wavelet Neural Network
N	Number of Balls
F	Rotating Frequency of Shaft
D_b	Ball Diameter
D_p	Pitch Diameter
θ	Contact Angle
$p(x)$	Probability Density Function
$\psi_r(t)$	Real and Even Wavelet
$\psi_i(t)$	Imaginary and Odd Wavelet
j	Number of The Stage of Decomposition
n	Peak Value of The Wavelet

$h_0(n)$	Low-Pass Filter for The Real Wavelet Transform
$h_1(n)$	High-Pass Filter for The Real Wavelet Transform
$g_0(n)$	Low-Pass Filter for The Imaginary Wavelet Transform
$g_1(n)$	High-Pass Filter for The Imaginary Wavelet Transform
ω_c	Center Frequency
BW	Bandwidth of The Center Frequency
$H(n, f)$	Complex Envelope of Signal at Frequency F
$\left\langle \lvert H(n, f)\rvert^4 \right\rangle$	Temporary Average of The Envelope Signal
$SK_Y(f)$	SK of Signal $y(t)$
$SK_X(f)$	SK of Signal $x(t)$
$SNR(f)$	Signal to Noise Ratio
K	Targeted Sparsity
a	Updated Atom
E	Error Matrix
$x(n; f, \Delta f)$	Vibration Signal of The Bearing At The Corresponding Frequency Band
$\bar{e}_x(f, \Delta f)$	Energy Fluctuation
$Var_x(f; \Delta f)$	Variance of The Energy Flow
$K_x(f; \Delta f)$	Spectral Kurtosis
$E_x(a; f, \Delta f)$	Square Envelope Spectrum
I	Number of Realizations
$\mathbf{k}$	Gain Vector
$\frac{da}{dN}$	Rate of The Crack Extension

J	Current Density
h_e	Equilibrium Gap Size
AIC	Akaike Information Criterion
AR	Auto Regressive
ARCH	Auto-Regressive Conditionally Heteroscedastic
ARIMA	Auto-Regressive Integrated Moving-Average
ARMA	Auto Regressive And Moving Average
BPFI	Ball Pass Frequency, Inner Race
BPF	Ball Pass Frequency
BPFO	Ball Pass Frequency, Outer Race
CBM	Condition-Based Maintenance
CCD	Charge Coupled Device
CEEMD	Complementary Ensemble Empirical Mode Decomposition
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DTWT	Dual-Tree Wavelet Transform
DWT	Discrete Wavelet Transform
ECG	Electrocardiogram
ECM	Electro-Chemical Machining
EDM	Electrical Discharge Machining
EEMD	Ensemble Empirical Mode Decomposition
EKF	Extended Kalman Filter
EMD	Empirical Mode Decomposition
EWT	Empirical Wavelet Transform

FFT	Fast Fourier Transform
FTF	Fundamental Train Frequency
GMM	Gaussian Mixture Model
HAZ	Heat-Affected Zone
HFRT	High Frequency Resonance
HMM	Hidden Markov Model
IMF	Intrinsic Mode Function
KPIs	Key Performance Indicators
K-SVD	K Means Singular Value Decomposition
LBM	Laser Beam Machining
LMS	Least Mean Square
LP	The Lundberg And Palmgren Model
MA	Moving Average
MAE	Mean Average Error
MAPE	Mean Average Percentage Error
MAP	Maximum A Posteriori
MCA	Morphological Component Analysis
ML	Machine Learning
MSE	Mean Square Error
NN	Neural Network
OMP	Orthogonal Matching Pursuit
PCA	Principal Component Analysis
ReLU	Rectified Linear Unit
RLS	Recursive Least Square
RMS	Root Mean Square

RUL	Remaining Useful Life
RV	Realized Volatility
SE	Square Envelope
SES	Square Envelope Spectrum
SK	Spectral Kurtosis
SNR	Signal-To-Noise Ratio
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVR	Support Vector Regression
TQWT	Tunable Q-Factor Wavelet Transform
TSA	Time Synchronous Averaging
UKF	Unscented Kalman Filter
VFF	Variable Forgetting Factor
WNN	Wavelet Neural Network

SUMMARY

As the technology in mechanical engineering advances, increasing complex mechanical systems are developed to achieve better quality and accuracy with prolonged useful life. However, with the added components and processes included in the mechanical system, the relationship between the input process parameter and output process parameter is difficult to define precisely. Unlike the simple spring-mass-damper system, which the displacement, velocity and acceleration can be described by linear differential equations, the system exhibits a high dimensional nonlinear relationship from the input parameter to the output parameter. Additionally, with more components included in the system, the probabilities of a machinery breakdown increase. Although the quality of the components improves, it does not offset the complication of the added complexity. Traditional scheduled maintenance cannot accommodate to the change of configurations of the system in a timely manner. Therefore, condition-based-maintenance (CBM) is implemented to reduce cost and improve reliability. The CBM monitors the actual running condition of the mechanical systems by using data acquired from different sensors. After the data acquisition, various signal processing techniques are implemented to denoise and extract critical features from the acquired data. The analysis of the data yields useful information for the prognostic algorithms to predicate the types of damage and the remaining useful life (RUL) of the system. Statistical methods are generally added in the prognosis to address the stochastic nature of the mechanical system and determine an interval for optimal maintenance. The addition of machine learning tools helps solving the complex problems without the requirement of complete understandings of the sophisticated systems

or analytical models. The major two categories machine learning are implemented to understand mechanical systems are regression and classification. In machinery diagnosis, the machine learning (ML) tools are mainly used to adjust model parameters to gain enhanced performance of noise reduction and fault location recognition. In machinery prognosis, the ML tool are mainly used to predict defect size and the RUL.

The contribution of this study is to implement intelligent algorithms with adaptation and various machine learning tools to assist the diagnosis of incipient fault signature of rotating machineries and propose a possible route for diagnose machinery remaining life. The heuristic approach refers to implementation of various computationally efficient algorithms to accelerate the model diagnosis and prognosis. More specifically, various signal processing techniques including wavelet decomposition, principal component analysis, envelope analysis, spectral kurtosis, dictionary learning, deep learning are implemented to enhance the fault signature covered in background noise. The optimization algorithms contain recursive least square, LASSO, unscented Kalman filter (UKF) and Bayesian Algorithm to update the diagnostic model parameters both linearly and nonlinearly. As the machine degradation occurs, the parameters of the diagnostic model need to be updated to accurately capture the degradation trend and fault location. The need for adaptation of the model parameters cannot be overemphasized. The development of the new approaches involves two major parts. The first part of this study focuses on various signal processing techniques to obtain critical features in the acquired signal from the mechanical system during the early degradation stage. The deep learning algorithms are examined to diagnose bearing degradation stage based on a previous empirical model to estimate defect size. Case studies using the vibrational signal of rolling element bearings

will be demonstrated to prove the validity and applicability of the developed algorithm. The second part of this study concentrates on prognosis of bearing degradation pattern and remaining life. The auto regressive and moving average (ARMA) model with variable forgetting factor recursive least square (VFF-RLS) prediction model is implemented to predict the short-term behavior of bearing vibration signals. The wavelet neural network is implemented to predict bearing remaining life.

The weighted least square algorithm is successfully implemented in the dictionary learning algorithm to update the dictionaries during the degradation of bearings. In comparison to the traditional dictionaries updating scheme, the proposed method significantly reduces the computation complexity, which facilitate the implementation for online monitoring. The UKF is implemented to update the dictionaries weights nonlinearly. The result also indicates a potential candidate to be used to adaptively adjust the model parameter. However, during the study, the UKF behaves unstably when the training data is limited. The effect of amount of training data should be investigated in the future studies. The Bayesian algorithm implemented in deep learning network shows great capability in search the optimal set of parameters to be used in complex neural network. The result is compared to manually tuned network and support vector machine (SVM). The efficiency of the parameter of Bayesian algorithm is the most efficient for models with larger parameter. The selection of parameter boundaries still requires further investigation because the specific search boundaries may not include the global extrema. From all the examined methods, the CEEMD preprocessing algorithm is the most promising method to extract bearing outer race fault signature. A clear degradation pattern is obtained after the CEEMD, and the processed signal can be used for both diagnosis and prognosis. The

drawbacks of the CEEMD is the processing time. The algorithm requires significant amount of computation to be able to separate different frequency band, which possibly prohibit the online application. In the future, a more effective decomposition method should be investigated.

The VFF-RLS parameter updating scheme demonstrates great capability in updating ARMA model parameters in this study. But the ARMA model is limited to short term prediction. The capability of prediction of the deep learning algorithm should be investigated in combination with the applied load, rotating speed and acquired vibration signal in the future.

The predication models and optimization models are not limited in the area in machine condition monitoring. The application of predicting output of an electrochemical drilling process is investigated. The prediction accuracy of deep convolutional network shows the best result in comparison to the traditional neural network and the developed physics based neural network. Although the prediction accuracy is on the order of less than 15%, to be implemented in the actual application, the algorithm will still need to be improved. The effect of adding more input parameters into the prediction model and expand the data size will significantly improve the prediction result.

CHAPTER 1. INTRODUCTION

1.1 ROTATING MACHINERY CONDITION MONITORING

Machine condition monitoring is a critical part to ensure safe and efficient production. The condition-based maintenance (CBM) is the most efficient maintenance strategy implemented across various industries. The traditional maintenance schedule cannot meet today's demand because occasionally catastrophic failure happens within machines' components with serious consequences [1]. The complex operating condition could result in early failure of machineries. The online monitoring scheme checks the machine operating conditions constantly, which significantly reduces the occurrence of catastrophic failures.

The condition monitoring relies on two main techniques to obtain the machines' conditions. The first one is the vibration analysis. The fault within the machine develops as the machine starts operating. By monitoring the vibration through accelerometer or acoustic emission sensors, the mechanical signature analysis [2] can be performed to identify defect severity and fault location. The second type of techniques is the lubricant analysis. The lubricant carries information from the machine in the form of wear particles, chemical contaminants, etc. The temperature of the lubricant could also indicate the health condition of the machines if the machines are constantly running.

The vibration monitoring is the most popular among all the available monitoring methods because of its low cost, ease of installation, and good correlation with machine degradation conditions. It also reacts immediately to change in comparison to lubricant

analysis. Another benefit of the vibration analysis is its capability to identify the fault location based on the theoretically calculated vibration frequencies of the components of rotating machines. The most important feature of the vibration analysis is its compatibility with various signal processing techniques. With appropriate processing algorithm selected, the early weak fault can be detected to better estimate the machine condition and predict the remaining life. In this study, the vibration signal from rolling element bearings is used to test the proposed algorithms.

1.2 DIAGNOSIS OF MACHINE FAILURE

A rolling element bearing normally consists of outer race, inner race, cage and ball as shown in Figure 1 [3]. Each individual component generates different vibrational frequency as the bearing operates.

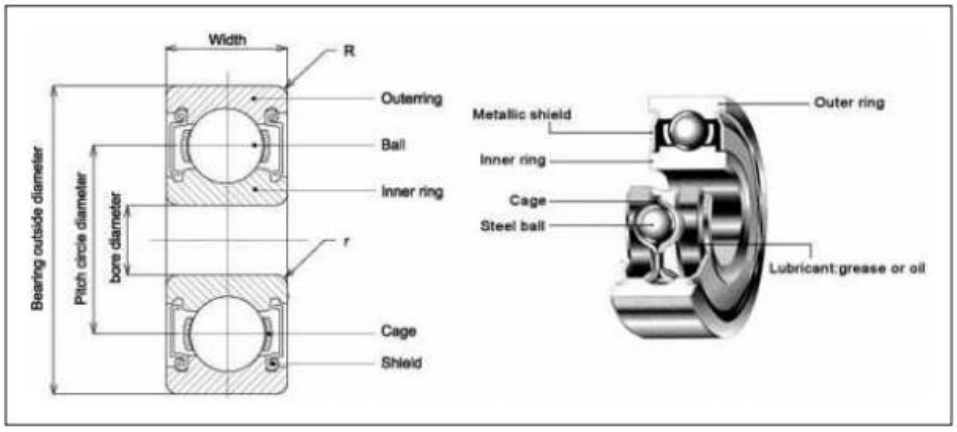


Figure 1 – Rolling Element Bearing [3]

The ball pass frequency of the inner race (BPFI), ball pass frequency of the outer race (BPFO), fundamental train frequency (FTF), and ball pass frequency are calculated by:

$$BPFI = \frac{N}{2} \times F \times (1 + \frac{D_b}{D_p} \times \cos\theta) \tag{1}$$

$$BPFO = \frac{N}{2} \times F \times (1 - \frac{D_b}{D_p} \times \cos\theta) \tag{2}$$

$$FTF = \frac{F}{2} \times (1 - \frac{D_b}{D_p} \times \cos\theta) \tag{3}$$

$$BPF = \frac{D_p}{2D_b} \times F \times [1 - (\frac{D_b}{D_p} \times \cos\theta)^2] \tag{4}$$

where N is the number of balls, F is the rotating frequency of shaft, D_b is the ball diameter, D_p is the pitch diameter and θ is the contact angle [4].

The general statistical terms used in diagnosis of rolling element bearings are the variance which correspond to the moment of inertia about the mean value defined in (5):

$$\sigma^2 = \int_{-\infty}^{\infty} [x - \mu]^2 \, p(x) dx \tag{5}$$

where x is the signal, μ is the mean value of the signal, $p(x)$ is the probability density function. The third centered moment and the fourth moment are called skewness and kurtosis respectively. They are denoted by (6) and (7):

$$S = \frac{\int_{-\infty}^{\infty} [x - \mu]^3 \, p(x) dx}{\sigma^3} \tag{6}$$

$$K = \frac{\int_{-\infty}^{\infty} [x-\mu]^4 \, p(x)dx}{\sigma^4} \tag{7}$$

The above parameters along with the root mean square value are the most widely used parameters for monitoring of degradation of rolling element bearings in time domain.

The frequency domain analysis includes the Fourier Analysis in discrete form and the Fast Fourier Transform for easy implementation. The Hilbert Transform and Demodulation are also used frequently for analysis the signal in frequency domain. The more popular analysis lies in the time-frequency domain. The short time Fourier Transform uses a windowing method to move along the overall signal and monitor the frequency change in each window. The Wigner-Ville Distribution are generally used for comparison of time-frequency distributions. The wavelet analysis implements different forms of wavelet to achieve signal denoising and reconstruction. The Spectral Correlation has great success in analyzing local bearing fault.

The general diagnostics approach includes data acquisition from measurement of sensors, data analysis using signal processing approach, and identification of the fault location and decision making [5]. The general used sensors are the temperature, debris, acceleration and acoustic emission sensors [6]. The temperature sensors are mounted on the surface of bearing housings to monitor the change of temperature. In general, an increase of temperature indicates the incipient failure of the bearings. However, no clear evidence is reported to relate the severity of damage to the change of temperature. Therefore, this method is not popular in most recent research. Another early approach of

measurement relies on measuring the amount of the debris of the bearings by using magnetic plugs in the lubricant. A trend of the amount of debris can be captured to detect the initiation of the failure of the bearing. However, this approach requires disassembling and reassembling the components of the bearings, which is time consuming and inefficient. The vibrational measurements are the most accepted indicator nowadays for diagnosis because of its good correlation with defect size and ease of implementation [7]. The piezoelectric accelerometers and the acoustic emission sensors are commonly used to obtain the vibration signals.

After the desired signal is acquired, different signal processing techniques are implemented to reduce the noise of the signal and extract the critical features. A decision is made based on the detected severity and the location of the fault after the extraction of the feature of the signal. Early signal processing techniques in the time domain such as using the skewness, kurtosis, crest factor are implemented to identify defect [8]. The defect generally increases with those indicators and a threshold value can be set to as the benchmark of performing maintenance. Other frequency domain methods such as the fast Fourier transform (FFT) and high frequency resonance techniques (HFRT) in combination with different mode decomposition methods are implemented to extract features in early documented research [9, 10]. The frequency methods are generally used to detect the fault location. However, those techniques cannot detect the defect presented in the system at an early stage because of the desired feature is heavily masked by the system noise. In addition, the FFT and HFRT can only extract features when the signal to noise ratio is large [11]. Signal processing techniques that can reduce the noise and extract critical features in the vibration signal are still under development.

Randall has proposed a general diagnosis for fault detection which involves deterministic and random signal separation, establishment of vibration criteria, implementation of time-frequency analysis to extract fault signature [5]. The current research for deterministic and random signal separation focuses on tacholess order tracking [12]. The goal is to reduce the effect of fluctuation of shaft rotating speed. However, the documented method, in general, requires small changes in shaft rotating speed. The most recent research proposed a method to accommodate large gradient in shaft speed [13]. The traditional vibration criteria or indicator includes the RMS, kurtosis, skewness. The most recent research uses the negentropy as an indicator to achieve prognosis because its sensitivity of repetitive transients and impulsiveness [14]. The new indicator could be a better choice than the traditional indicators in early fault detection.

1.3 PROGNOSIS OF MACHINE FAILURE

The prognosis of machine failure relies on the determination of RUL. The RUL are generally predicted based on physics-based and data-driven techniques [1]. The physics-based model requires a physical model such as the Paris' law for crack growth to estimate the degradation trend. The data-driven model is based on statistical analysis of measurement data from a database which contains a significant amount of experimental run-to-failure data. In the recent years, the hybrid models combining the physics and the data were developed to meet the industry requirement. Various machine learning algorithms are implemented to predict the defect growth and estimate the RUL. For instance, by measuring the duration of a short impulse signal created by repetitive contact with the defect site from the rolling element, a rough estimation of defect size can be obtained. The challenge lies in extracting a clear impulse signal from the acquired signal.

This task is extremely difficult to achieve in the early weak fault stage but relatively easy in the late degradation stage.

CHAPTER 2. DIAGNOSIS OF EARLY MACHINERY FAILURE

2.1 DIAGNOSIS OF ROLLING ELEMENT BEARING USING COMPLEX WAVELET

2.1.1 Introduction

This section presents a method using the Dual-Tree Wavelet Transform (DTWT) combined with an optimized Morphological Component Analysis (MCA) to extract the weak fault in the early degradation stage of bearings. The DTWT decomposes the signal into multiple layers based on the fundamental frequencies of bearings. The MCA takes the decomposed signal and separates the signal into two components as output. Alternating parameter selection is implemented to improve the result of the MCA. The weak fault signature of the bearing is extracted from the separated components of the MCA. Simulated and experimental data are used to validate this method. The proposed optimization method is compared with an unscented Kalman filter parameter optimization process. The proposed diagnostic model demonstrates the superior capability of the early detection of the faulty signal within bearings in comparison with the traditionally used wavelet decomposition method and the unscented Kalman filter.

One of the most widely used denoising techniques is the wavelet transform based denoising methods [15]. The wavelet transform shows good result in denoising and finding the band-pass frequency of the outer race [16]. Rubini and Meneghetti [17] points out the advantage of wavelet transforms in identifying the fault in bearings even if the surface is flattened because of contact stress. Prabhakar *et al.* [18] demonstrate the superiority of the

discrete wavelet transform (DWT) in detecting both single and multiple faults in bearings. However, the traditional wavelet transform has problems of singularities, shift variance, aliasing and lack of directionality [19, 20]. Because of using only one mother wavelet in traditional wavelet transform, the reconstructed signal is inevitably distorted. A solution to alleviate the problem using the complex wavelets was proposed in [19]. The dual-tree complex wavelet transform (DTWT) uses the real and imaginary parts of the wavelet coefficients, which forms orthonormal bases to overcome the issues in traditionally used wavelet transforms [19].

The morphological component analysis (MCA) based on sparse representation of signals proposed by Starck *et al.* demonstrates great outcomes in image processing [21]. The method is capable of separating images into texture and piecewise parts by using two different dictionaries. Chaibi *et al.* implements MCA with tunable Q-factor wavelet transform (TQWT) to distinguish the transient with and without high frequency oscillations. The result obtained by using the TQWT-MCA reduces the false discovery rate from 30% to 7% [22]. Lajnef *et al.* uses a similar method to separate the component of sleep electroencephalography signals with the improvement of the false discovery rate [23]. Imani *et al.* implement MCA to extract the retinal blood vessels in combination with a traditional wavelet transform [24].

The DTWT and MCA demonstrate significant advantages in multi-dimensional data analysis separately. However, to the best of our knowledge, no documented research attempt to combine the benefits of the two methods for bearing diagnosis. This section presents a bearing diagnostics model by implementing a combination of the DTWT and improved MCA to overcome the difficulties in detecting the weak fault signature. The

DTWT is used to decompose the vibration signal into the desirable range of frequency while reducing the unwanted system noise. Principal component analysis (PCA) is implemented to choose the layer of the decomposition signal which contains the most useful information [25]. The MCA further separate the layer of decomposed signal with the highest PCA score into the fault signal generated by a single fault presented in the bearing and other frequencies. An iterative alternating parameter selection algorithm is implemented to optimize the separation result of the MCA. Hilbert transform is used to show that the fault frequency is detected early in the degradation of the bearing. Simulated and experimental data are used to validate the proposed method. The newly developed diagnostic model detects the weak fault signal in the early degradation stage of the bearing. The model is compared with a traditional wavelet transform with Hilbert transform to demonstrate its superiority in detecting weak fault signature in the early degradation stage.

2.1.2 Diagnostic Model

The proposed model in this section is shown in Figure 2. The early vibration signal obtained by an accelerometer is passed to the DTWT for denoising and decomposition. PCA is implemented to select the signal which contains the most useful information for diagnosis. The signal that contains the first principal component direction is selected. Note that the selected signal explains the most part of the variance of the vibration signal, thus, contains the most useful information. The selected signal is then fed into the MCA. The MCA further separates the bearing signal into the desirable fault signal and noises. The Hilbert Transform is used to reveal the characteristic frequency of the impulse signal. Based on the fundamental equation of calculating the bearing characteristic frequencies

[24], the fault location is identified. A decision is made based on the intensity and the frequency of the fault signal.

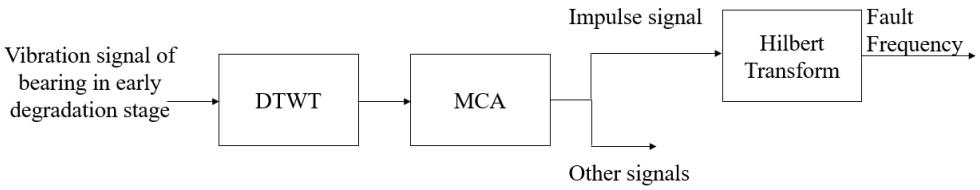


Figure 2 – DTWT+MCA Diagnostic Model

The fundamental fault frequencies are denoted by ball pass frequency inner race (BPFI), ball pass frequency outer race (BPFO), fundamental train frequency (FTF), and ball pass frequency (BPF). In general, the frequency or frequencies extracted using different signal processing method are compared with the frequencies listed above to locate the defect location. Early research [26] reveals that the intensity of the vibration signal is related to the defect size of the bearing. More recent research draws a similar conclusion about this relationship [27, 28]. In actual application, multiple factors need to be considered because the absolute magnitudes of the extracted signal at the fault frequencies are based on the methods used. For instance, the wavelet decomposition will significantly affect the magnitude of the original vibration signal. The experiment is normally required to correlate the measured signal and the defect severity.

2.1.2.1 DTWT

The wavelet transform proposed by Mallat [29] has numerous applications in the field of mechanical engineering for data compression, decomposition and signal denoising. An easy way to understand the wavelet transform is to imagine the signal is composed of

different wavelets. The wavelets are stretched and shifted to different parts of the signal to reconstruct the original signal. Although the wavelet transform has many favorable characteristics, it remains several problems in certain applications [19]. The first problem is the oscillations of the wavelet coefficients. The wavelet coefficients could oscillate between positive and negative around an impulse signal. To address this issue, Choi *et al.* proposed the model of complex wavelet transform to alleviate the problem of the oscillation of the coefficients and the overlap of a singularity [30]. The second problem is the shift variance. The wavelet transform suffers from the shift of the signal unless the shift is a multiple of two [31]. The third issue is aliasing. Because the wavelet transform is equivalent to implementing low-pass and high-pass filters, the aliasing of the signal occurs during decomposition. The aliasing leads to the inaccuracy in the reconstructed signal. For multi-dimensional processing, the lack of directionality of the traditional wavelet could result in difficulty of processing edges and ridges in geometric images [19].

These problems are alleviated by using the recently developed dual-tree complex wavelets [19]. Researchers noticed the Fourier Transform does not experience the same problems as for the wavelet transform because of the added complex oscillating part. By using the analogy of the Fourier Transform, an imaginary wavelet is added into the traditional wavelet to construct the complex wavelet as shown in (8):

$$\psi_c(t) = \psi_r(t) + j\psi_i(t) \tag{8}$$

where $\psi_r(t)$ is the real and even wavelet, and $\psi_i(t)$ is the imaginary and odd wavelet. The complex wavelet coefficient is shown in (9):

$$C_c(j,n) = C_r(j,n) + jC_i(j,n) \tag{9}$$

where *j* is the number of the stage of decomposition, *n* indicates where the peak value of the wavelet is, C_r and C_i are the real and imaginary coefficients.

The schematic diagram for the dual tree wavelet transform is very similar to the discrete wavelet transform. It has the analysis (decomposition) and the synthesis (reconstruction) parts. The schematic for the analysis and synthesis part of the dual-tree wavelet is shown in Figure 2 and 3. Because of the dyadic nature of the DTWT, the decomposition level stops once the frequency band of the signal cannot be further split in half.

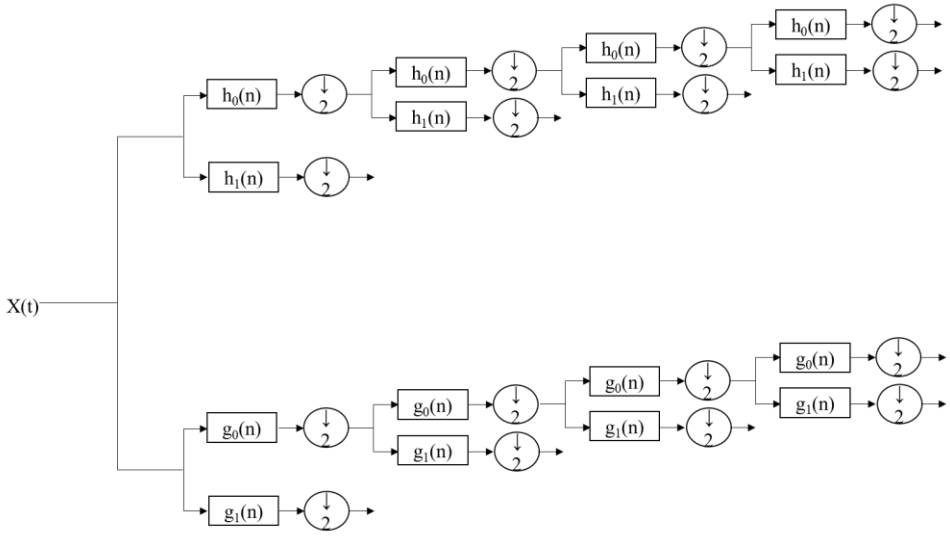


Figure 3 – Analysis of DTWT [19]

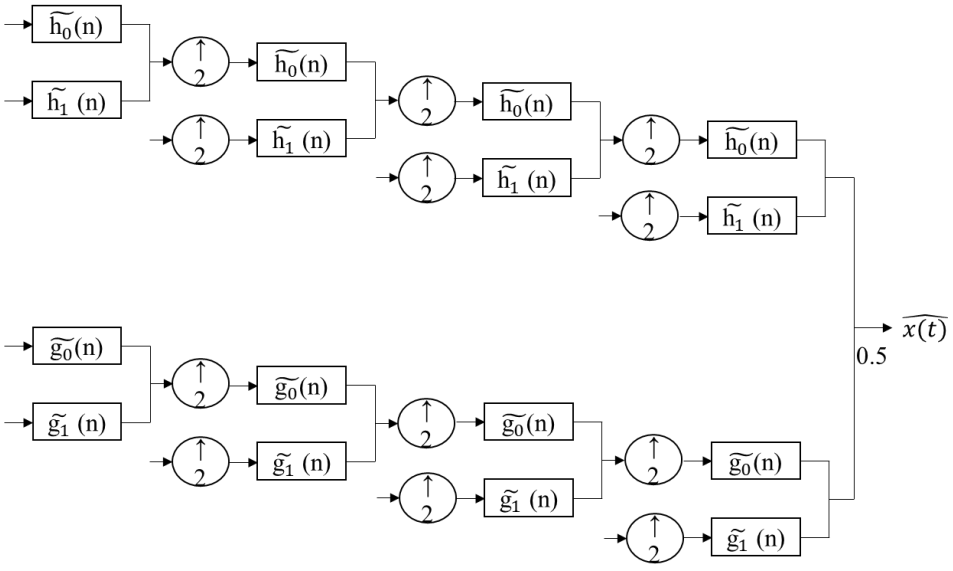


Figure 4 – Synthesis of DTWT [19]

The $h_0(n)$ and $h_1(n)$ are the low-pass and high-pass filter pair for the real wavelet transform $\psi_r(t)$. The $g_0(n)$ and $g_1(n)$ are the low-pass and high-pass filter pair for the imaginary wavelet transform $\psi_i(t)$. $\psi_i(t)$ is selected to be the Hilbert transform of $\psi_r(t)$.

The dual-tree complex wavelet is implemented as the basis for wavelet decomposition. The level of decomposition is selected to be 7 because of the sampling frequency and the frequencies of interest. The first-stage Farras filters and Kingsbury Q-shift filters for subsequent stages are implemented. The taps of the Kingsbury Q-shift filters are selected to be 14 by trial and error.

Principal component analysis (PCA) is implemented after the DTWT. The general purpose of implementing PCA is to extract the most information from data [32]. PCA computes the principal components which are linear combinations of the original data and decompose them into several orthogonal directions. The first principal component explains the largest variance and the second has the largest possible inertia [32]. The principal components are obtained from singular value decomposition of the data. The singular value decomposition of the data matrix X is:

$$X = P\Delta Q^T \tag{10}$$

where P is the $I \times L$ matrix of left singular vectors, Δ is the diagonal matrix of singular values, and Q is the $J \times L$ matrix of right singular vectors. The PC score F is generally represented by (11):

$$F = P\Delta \tag{11}$$

By rearrange terms from (4) and (5), the score F can be calculated by (12):

$$F = XQ \tag{12}$$

The decomposed layer with the most information can be found using the PCA by adapting the signal most related to the first principal component.

2.1.2.2 MCA

Morphological Component Analysis is a sparse-based innovative decomposition method aiming to separate signal into two distinctive parts: texture and piecewise. This method assumes that for every signal to be separated, a dictionary that enables its construction using sparse representation exists [33]. In addition, the assumption is based on the condition that the two different dictionaries are independent and only works for its own signal behavior. For instance, a signal X_t of length N consists of only the pure texture content. There exist an overcomplete representation matrix T_t which enable a sparse solution of (7):

$$\underline{\alpha}_t^{opt} = Arg \min_{\underline{\alpha}_t} \left\| \underline{\alpha}_t \right\|_0 \text{ subject to: } \underline{X}_t = T_t \underline{\alpha}_t \tag{13}$$

where $\left\| \underline{\alpha}_t \right\|_0$ is the L_0 norm of the vector $\underline{\alpha}_t$. T_t is under the assumption that sparse representation is valid if the texture appears in parts of X_t and otherwise zero. The overcomplete transform of the signal X_t is achieved with a representation $\underline{\alpha}_t$ which maximize the sparsity.

The forward transform with T_t applied to the signal consists of no texture and pure piecewise signature should yield non-sparse representations. The major function of the dictionary T_t is to distinguish between the texture and piecewise signature and always

choose the texture over the natural part. A possible indicator of fidelity of the dictionary is the functional shown from (14) to (16):

$$T_t^{opt} = Arg \min_{T_t} \frac{\sum_k \left\| \underline{a}_t^{opt}(k) \right\|_0}{\sum_j \left\| \underline{a}_t^{opt}(j) \right\|_0} \tag{14}$$

$$\underline{a}_t^{opt}(k) = Arg \min_{\underline{a}_t} \left\| \underline{a}_t \right\|_0 \text{ subject to: } \underline{X}_t(k) = T_t \underline{\alpha}_t, k = 1,2,3... \tag{15}$$

$$\underline{a}_n^{opt}(j) = Arg \min_{\underline{a}_n} \left\| \underline{a}_n \right\|_0 \text{ subject to: } \underline{X}_n(j) = T_t \underline{\alpha}_n, j = 1,2,3... \tag{16}$$

Similarly, a different dictionary T_n can be implemented so that it always prefers the piecewise content versus the texture. Therefore, for an arbitrary signal $\underline{X}$ consists of both texture and piecewise content, the sparsest representation over the combined dictionary of T_t and T_n should be the goal of the MCA. To achieve this, (17) need to be solved:

$$\{\underline{a}_t^{opt}, \underline{a}_n^{opt}\} = Arg \min_{\{\underline{a}_t, \underline{a}_n\}} \left\| \underline{a}_t \right\|_0 + \left\| \underline{a}_n \right\|_0 \text{ subject to: } \underline{X} = T_t \underline{\alpha}_t + T_n \underline{\alpha}_n \tag{17}$$

where $T_t \underline{\alpha}_t$ is the texture component of the signal and $T_n \underline{\alpha}_n$ is the piecewise component of the signal. The success of the separation depends on the selection of T_t and T_n. In practice, L_0 norm requires to solve a non-convex problem which is NP hard [34]. Therefore, L_0 norm are generally replaced by an L_1 norm which leads to an optimization problem with solution. The L_1 norm is represented in (18):

$$\{\underline{a}_t^{opt}, \underline{a}_n^{opt}\} = Arg \min_{\{\underline{a}_t, \underline{a}_n\}} \|\underline{a}_t\|_1 + \|\underline{a}_n\|_1 \text{ subject to: } \underline{X} = T_t \underline{\alpha}_t + T_n \underline{\alpha}_n \quad (18)$$

The selection of the dictionary for the MCA does not have a specific guideline. There exists a wide variety of dictionaries that potentially work for separating the impulse signal from the noisy signal of bearings. The Tunable Q-factor Wavelet Transform (TQWT) has demonstrate the advantage over the traditional wavelet transform [35-38]. It decomposes the signal into low-pass sub-band and high-pass sub-band signal with sampling rates of αf_s and βf_s. The variable f_s represents the sampling frequency of the input signal. The value of α and β satisfy the condition [38]:

$$0 < \alpha < 1, 0 < \beta \leq 1 \quad (19)$$

The perfectly separated signal should have value at its own frequency sub-band and equal to zero elsewhere. To ensure the good localization of the filter response, α and β need to satisfy [38]:

$$\alpha + \beta > 1 \quad (20)$$

The desirable signal is separated by using the tunable Q-factor related to resonance property of the signal as shown in (21):

$$Q = \frac{\omega_c}{BW} \quad (21)$$

where ω_c is the center frequency, and *BW* is the bandwidth of the center frequency. The other parameters related to the TQWT algorithm is the redundancy, *r* and the number of levels of the TQWT, *J*. The redundancy is the over-sampling rate when the TQWT is implemented and it is shown in (22):

$$r = \frac{\beta}{1-\alpha} \tag{22}$$

The sampling rate at the sub-band *j* is $\beta\alpha^{j-1} f_s$ and the summation of all the sub-bands *j* yields the sampling rate of $\beta/(1-\alpha) f_s$. The *j*-level frequency response in the interval (ω_1, ω_2) is:

$$\omega_1 = (1-\beta)\alpha^{j-1}\pi, \omega_2 = \alpha^{j-1}\pi \tag{23}$$

The center frequency is defined as:

$$\omega_c = \frac{1}{2}(\omega_1 + \omega_2) = \alpha^j \frac{2-\beta}{2\alpha}\pi \tag{24}$$

$$f_c = \alpha^j \frac{2-\beta}{4\alpha} f_s \tag{25}$$

The relationship between α, β, j and *Q* is shown in (26):

$$\beta = \frac{2}{Q+1}, \alpha = 1 - \frac{\beta}{r} \tag{26}$$

The general rule of choosing those parameters is: $Q \geq 1, r \geq 3$ [39]. Selesnick also proposed some guidelines to optimize the TQWT algorithm in [40]. Under the assumption that signal *x(t)* consists of $x_1(t)$ and $x_2(t)$ can be sparsely represented in bases s_1 and s_2, the decomposed signal $x_1(t)$ and $x_2(t)$ can be found by minimizing the objective function with respect to w_1 and w_2:

$$J(w_1, w_2) = \left\| x - s_1 w_1 - s_2 w_2 \right\|_2^2 + \lambda_1 \left\| w_1 \right\|_1 + \lambda_2 \left\| w_2 \right\|_1 \tag{27}$$

The estimated signal after separation are denoted by:

$$x_1 = s_1 w_1, x_2 = s_2 w_2 \tag{28}$$

The selection of *Q* for different filter banks should follow:

$$\rho_{\max}(Q_1, Q_2) = \sqrt{\frac{Q_1 + 0.5}{Q_2 + 0.5}} \ll 1, Q_2 > Q_1 \tag{29}$$

where $\rho_{\max}(Q_1, Q_2)$ is the coherence. By adding this constraint, an alternating optimal parameter selection algorithm is implemented to selection the best value of *Q*, λ_1 and λ_2, which are the maximizer of the selection criterion established on (30):

$$Ratio = 0.3 \times \frac{M_{Q1,f}}{M_{Q2,f}} + 0.3 \times \frac{M_{Q1,2f}}{M_{Q2,2f}} + 0.2 \times \frac{M_{Q1,3f}}{M_{Q2,3f}} - 0.2 \times \frac{M_{Q1,s}}{M_{Q2,s}} \tag{30}$$

where $M_{Q1,f}$ is the magnitude of the signal at the defect frequency from the component separated using Q_1, $M_{Q2,f}$ is the magnitude of the signal at the defect frequency from the component separated using Q_2. The subscript *2f* and *3f* represent the harmonics of the defect frequency. The subscript *s* represents the shaft frequency. The optimized MCA parameters are found by maximizing this ratio of the two separated components from MCA.

The iterative alternating optimal parameter selection algorithm is shown in Figure 4. The detailed optimal selection procedure is described as following. Given the initial values of Q_2, λ_1 and λ_2, the optimal value of Q_1, which is denoted as $\hat{Q}_1$, is calculated to maximize the defect signal to noise ratio defined in (24). Then given $\hat{Q}_1$ and initial value of λ_1 and λ_2, a similar search strategy is used to find the optimal value of Q_2, $\hat{Q}_2$, corresponding to the maximum ratio in (24). The searching process is iterated until the value of $\hat{Q}_1$ and $\hat{Q}_2$ are converged, i.e. the change of defect signal ratio is less than 5% for any consecutive two iterations. After Q_1 and Q_2 are determined, λ_1 and λ_2 are located using the same methodology. The algorithm then checks the optimal values of Q_1 and Q_2 again to make sure it satisfies the 5% of change criterion. The alternatively determination the optimal value of Q_1, Q_2, λ_1 and λ_2 is computationally efficient and robust to select the optimal values of parameters.

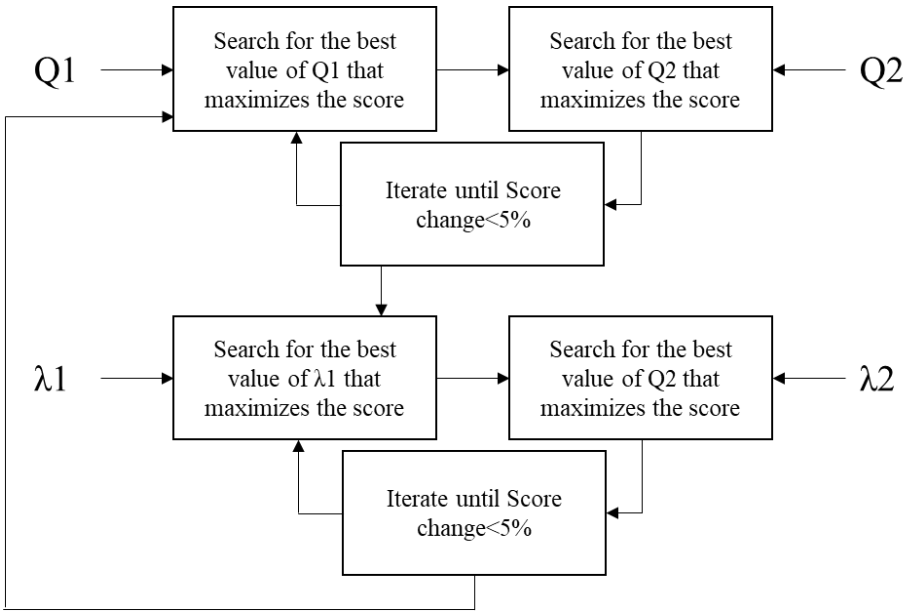


Figure 5 – Iterative Alternating Parameter Selection

The integration of the DTWT and improved MCA combined the advantage of the denoising from the dual tree wavelet and the separation capability from the MCA. It improves the separation of impulse signal from the previous documented research and provide a framework to optimize the parameter selection for signal processing by establishing the criterion of the separation of signal. In addition, the automatic iterative parameter selection helps reduce the tuning time for the MCA parameter rather than using an exhaustive searching method. The early weak fault signal presented in the bearing vibration signal can be extracted using this proposed method.

2.1.3 Simulation Validation

To validate the proposed method, a simulated signal is tested using the DTWT+MCA method. The simulated signal is composed of an impulse signal with 128 Hz frequency and

two other periodical signals with 20 Hz and 30 Hz frequency. The sampling rate is set as 2048 Hz. The simulated signal is created using (31) to (33):

$$s1 = 1.5(\sin(2\pi f_1 t) + \cos(2\pi f_2 t)) \tag{31}$$

$$s2 = 10\exp(-0.1t) \times \sin(2\pi f_3 \sqrt{1-(0.1)^2}\, t) \tag{32}$$

$$s3 = 6 \times whitenoise(t, 1, -1) \tag{33}$$

where *s1* is the signal with 20 Hz and 30 Hz frequency shown in Figure 6, *s2* is the signal with the impulse frequency of 128 Hz shown in Figure 7 and *s3* is the white noise shown in Figure 8. The combined signal is shown in Figure 9.

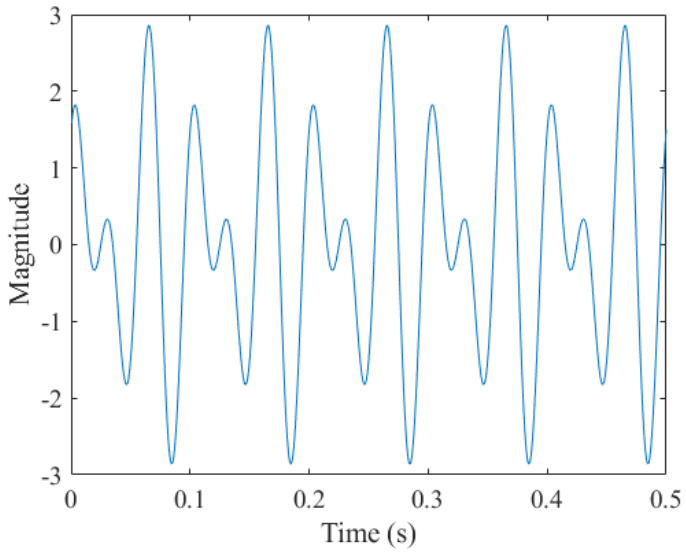


Figure 6 – Simulated Pure Sine with 20 Hz Frequency and Cosine with 30Hz Signal from Equation (31)

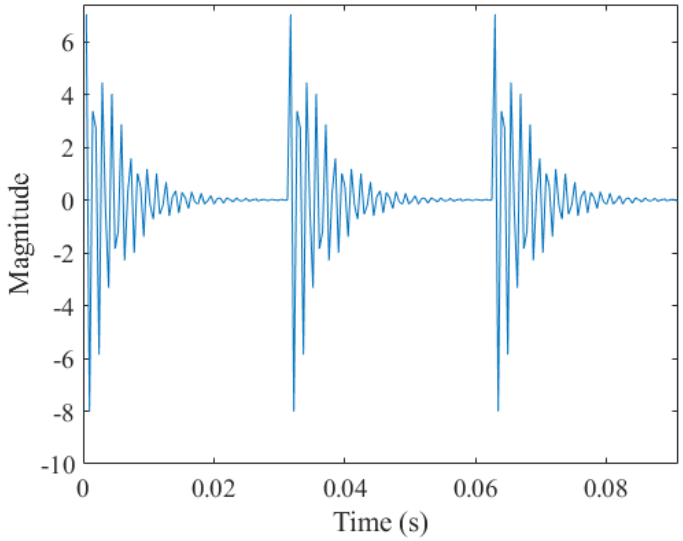


Figure 7 – Simulated Bearing Signal with 128 Hz Fault Frequency from Equation (32)

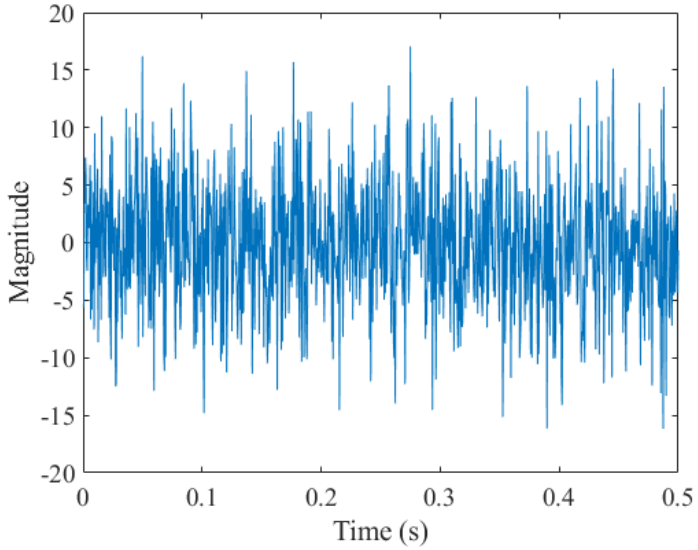


Figure 8 – Simulated White Gaussian Noise Signal from Equation (33)

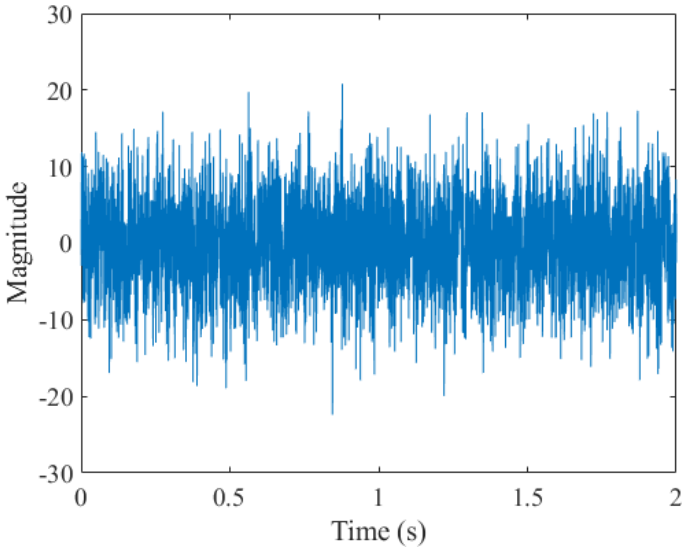


Figure 9 – Combined Simulated Signal with 20 Hz and 30 Hz Carrier Frequency, 128 Hz Fault Frequency and Added White Gaussian Noise

It can be seen from Figure 6 to 9 that the impulse frequency of the bearing is covered by the white noise, which simulates the early degradation stage of bearings. The Fast Fourier Transform of the signal is shown in Figure 10.

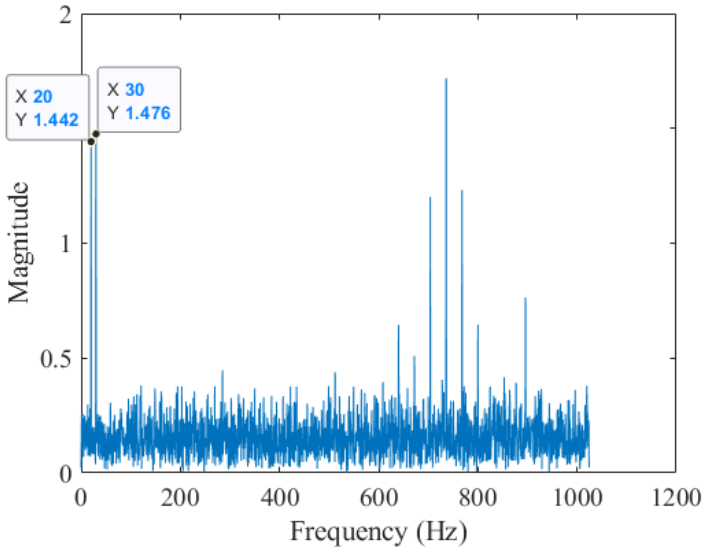


Figure 10 – Fast Fourier Transform of Simulated Signal

The impulse frequency is completely masked by the noise in the system. The proposed method is applied to the simulated signal. We decomposed the signal into 4 levels based on the sampling rate of 2048 Hz and desired impulse frequency of 128 Hz. The reconstructed signal by the DTWT is passed to the MCA for the separation of the impulse signal. Figure 11 shows the separated signals after the DTWT+MCA algorithm.

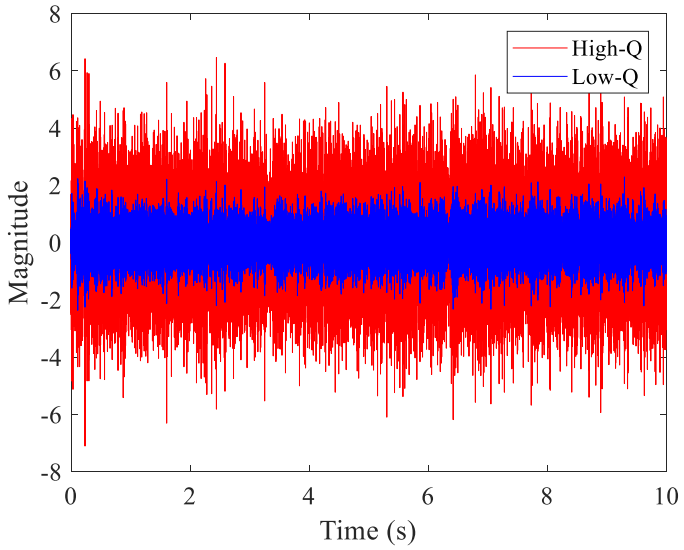


Figure 11 – Separated Signals

Using Hilbert Transform for the High-Q signal shown in Figure 11, the desirable 128 Hz signal is located under the heavy system noise as shown in Figure 12. In the next section, experimental data of the bearings are used to further validate this method.

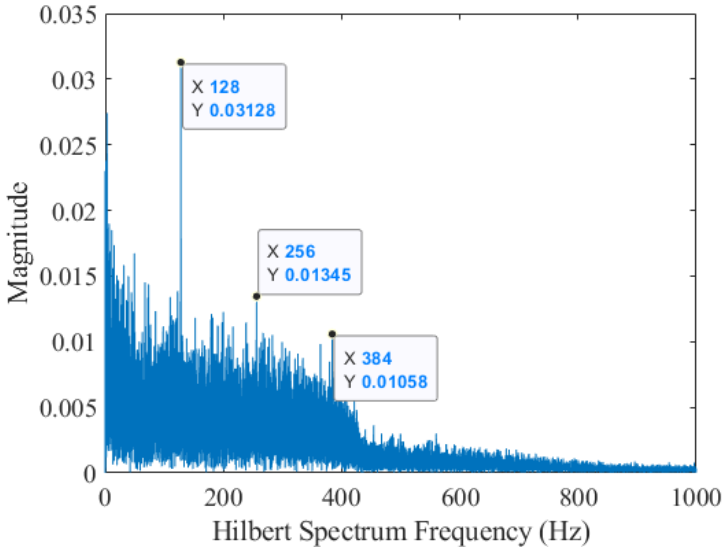


Figure 12 – Hilbert Transform for Impulse Signal

2.1.4 Experimental Validation

The experimental data of the vibration data of the bearings are obtained from Qiu *et al*.[16]. The experimental setup is shown in Figure 13. The test rig consists of four tests Rexnord ZA-2115 bearings. The rotating speed of the shaft is 2000 rpm and a radial load of 6000 lbs is applied to the shaft and bearing. The bearings are well-lubricated, and a magnetic plug is used to extract the debris in the lubricant. The stopping criteria of the test is when the amount of the debris which the magnetic plug collects exceed a preset value. A PCB 353B33 accelerometer was attached to each bearing housing and four thermocouples were installed on the outer race of each bearing. The vibration data was collected every 10 minutes using a NI DAQ 6062E.

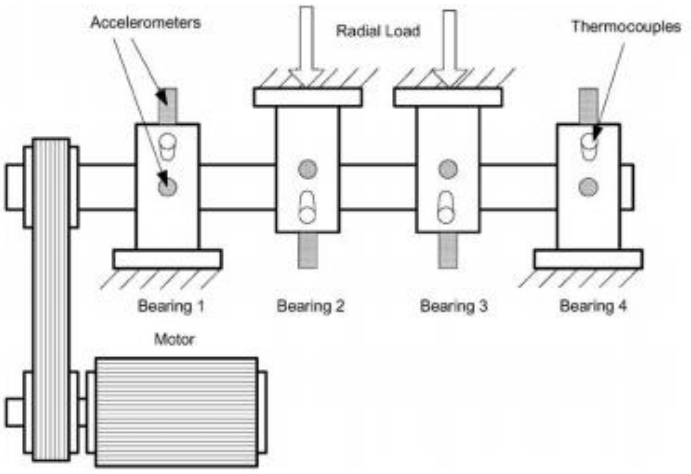


Figure 13 – The Test Rig of Bearing Vibration [16]

The vibration data of the bearing 1 with failure of its outer race is used to test the DTWT+MCA algorithm. Based on (8), the fault frequency for this bearing is calculated to be 236.4 Hz. The overall testing data is shown in Figure 14.

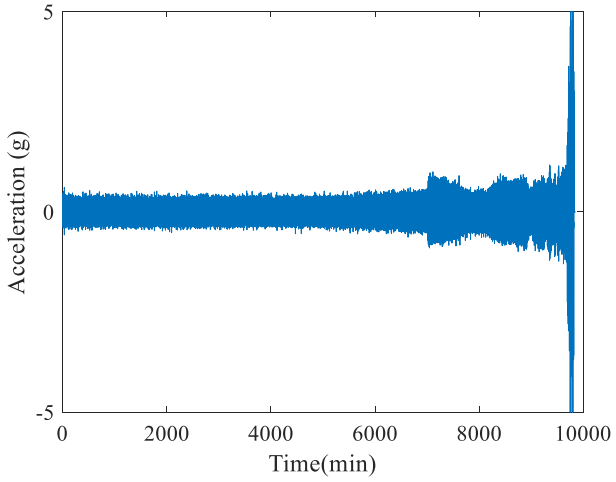


Figure 14 – Bearing Lifetime Test Data

The distinctive fault signal starts around 7000 minutes during the testing of the bearing. However, the actual initiation of the defect occurs earlier because the crack initiates in the sub-surface of the bearing rolling element which is hard to detect in the early stage. For the

purpose of early detection, the vibration data around 5350 minutes is used for weak fault detection. The vibration signal is shown in Figure 15. The impulse signal of the underlying fault at this stage of the degradation process is not very clear.

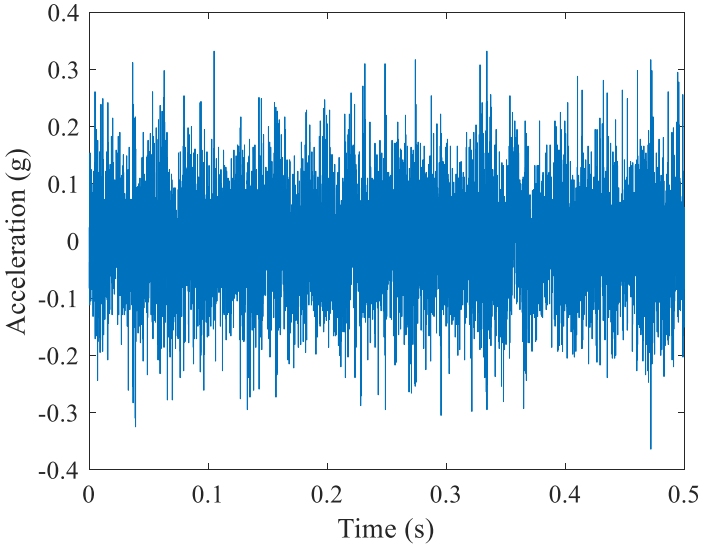


Figure 15 – Bearing Early Fault Signal

The DTWT first decomposes and reconstructs the original signal. The number of the decomposed is selected to be seven based on the sampling frequency of 2 kHz and fault frequency of 236.4 Hz. PCA is implemented to find the decomposed layer with the largest variance. The algorithm pursues the maximization of the variance as the maximum variance carries the most information of the data. The distribution of data corresponding to the decomposed layers with the largest variance is shown in Figure 16 with the variance value shown in Table 1.

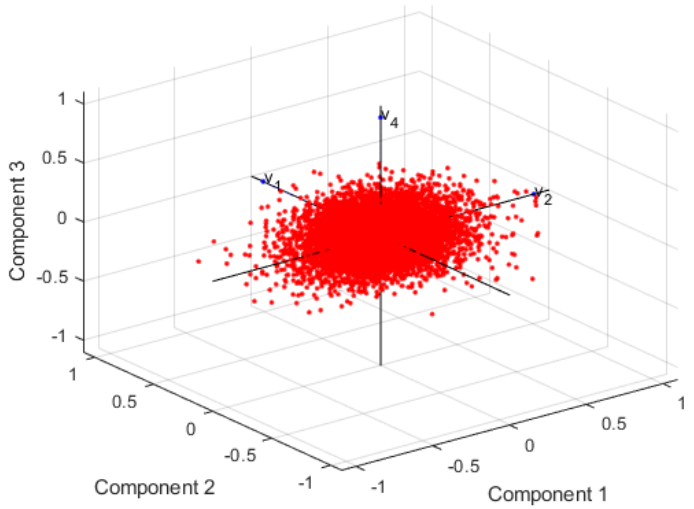


Figure 16 – PCA for Wavelet Decomposed Signal

It can be observed from Figure 16 that the three components containing the most information are the first, second and fourth layers of the wavelet decomposed signal. The spread of the data for the second layer is the largest as it has the largest variance.

Table 1 – Variance of Decomposed Signal from DTWT

Layer	Variance
1	3.7771×10^{-4}
2	7.3954×10^{-4}
3	1.9267×10^{-4}

Layer	Variance
4	3.0837×10^{-4}
5	4.1366×10^{-5}
6	1.3693×10^{-5}
7	6.4264×10^{-6}

The second layer of the decomposed signal is used for further processing because the information it contains based on the result of the PCA. The separation part of proposed DTWT+MCA method is applied to the signal at this time. Based on the sampling frequency of 20 kHz and the fault frequency of 236 Hz, a six-layer TQWT is used. The PCA is implemented again and the sixth layer of the decomposed signal is identified to have the most of useful information. This signal is fed into the MCA for impulse signal separation. The parameters of the MCA are selected using an iterative alternating parameter selection method with constrains based on [41]:

$$3 < Q_1 < 15, 1 < Q_2 < 2, r > 3, J > 1 \tag{34}$$

where Q_1 and Q_2 are the tunable Qs for the TQWT, r is the redundancy and J is the level of decomposition. To achieve the desirable separation of signal, the criterion defined in (24) is used with the value of the faulty signal frequency at 236 Hz and the faulty harmonics

at 472 Hz and 708 Hz. The values of Q_1, Q_2, λ_1, λ_2 are found to be 6.525, 1.01, 0.00125 and 0.04 respectively. The effect of the values of *r* and *J* on the ratio defined in (24) is negligible based on the test performed. The ratio between the separated signal is 650.5 based on the criterion and parameter selected. The separated signals are shown in Figure 17 and the Hilbert Transform of the High-Q component is shown in Figure 18.

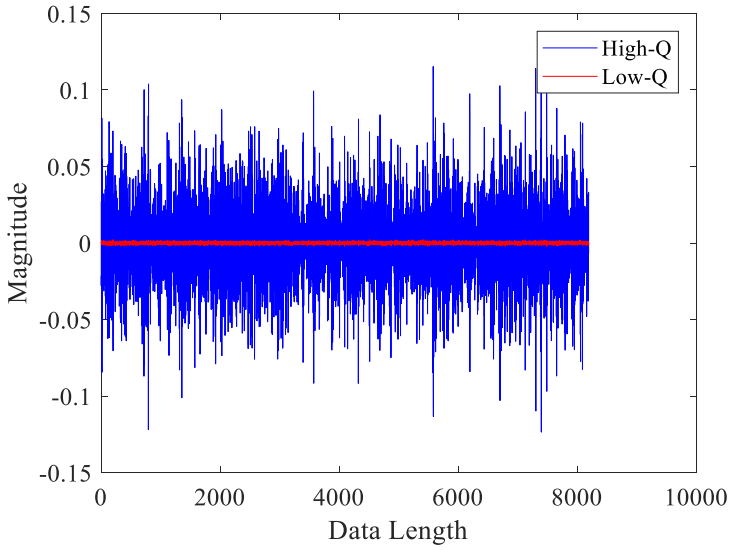


Figure 17 – MCA Decomposed Signals

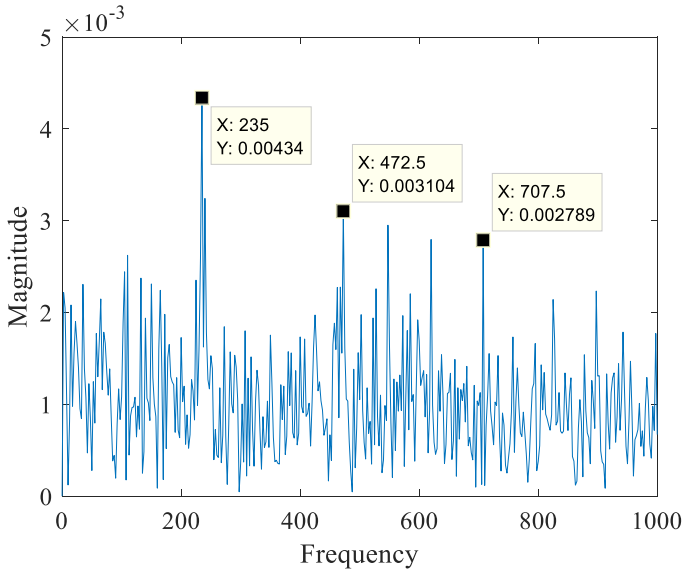


Figure 18 – Weak Fault Frequency by DTWT+MCA

It can be observed that the outer race fault frequency and its harmonic frequency can be detected in the early degradation of the bearing. A comparison of the separated High-Q component is shown in Figure 19 along with its signal to noise ratio shown in Table 2 for the untuned MCA separation result.

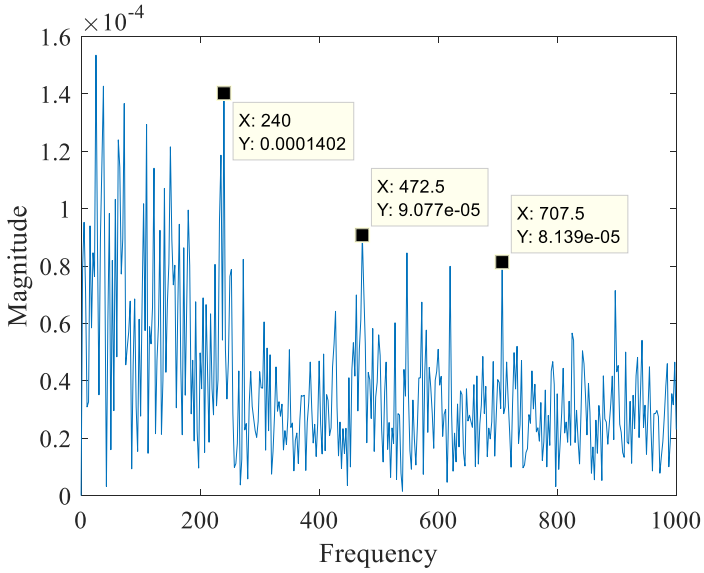


Figure 19 – Weak Fault Frequency by DTWT+MCA

Using the untuned parameters, the separation result shown in Figure 19 is not as desirable as shown in Figure 18. The signal to noise ratio is defined as:

$$SNR = \frac{M_1 + M_2 + M_3}{M_{index \neq 1,2,3}} \tag{35}$$

where M_1 is the magnitude of the signal at the fault frequency, and M_2, M_3 are the magnitude of the signal at the harmonics. $M_{index \neq 1,2,3}$ stands for the magnitude of the other frequencies. The calculated signal to noise ratio is shown in Table 2.

Table 2 – Comparison of Signal to Noise Ratio

Tuning Paremeter	Signal to Noise Ratio	Peak Signal to Noise Ratio	Magnitude at Fault Frequency
Q_1=6.525, Q_2=1.01, λ_1=0.00125, λ_2=0.04	0.0256	65.78	0.0043
Q_1=4, Q_2=2, λ_1=0.001, λ_2=0.001	0.0256	37.58	0.0021
Q_1=8, Q_2=1.5, λ_1=0.01, λ_2=0.01	0.0225	36.79	0.0015
Q_1=12, Q_2=1.8, λ_1=0.1, λ_2=0.3	0.0255	35.52	0.0017

It can be observed from Table 2 that the iterative alternating parameter selection generates a separated signal with higher signal to noise ratio and magnitude of signal at the fault frequency than using values defined within the range from [41].

The generally used wavelet transform with Hilbert transform and the dual-tree discrete wavelet are used as a benchmark model to compare the performance with the

proposed model. The wavelet uses the Daubechies wavelet and the decomposition level is set to be the same as the DTWT. The extracted frequency of the wavelet transform and Hilbert transform is shown in Figure 20. Although the fault frequency at 236 Hz can be seen from Figure 20, the signal to noise ratio of 0.0095 is relatively low in comparison to the result shown in Table 2. In addition, the harmonics in Figure 20 are not as distinctive as shown in Figure 18. Because of the nature of the dual-tree discrete wavelet transformation, the signal is critically sampled versus oversampled for the dual-tree complex wavelet transform. It can be observed from Figure 21 that significant diagnostic information is lost during the discrete wavelet transformation. Therefore, the proposed method yields a better result of the relative magnitude of the desirable fault frequencies over the other noise frequencies.

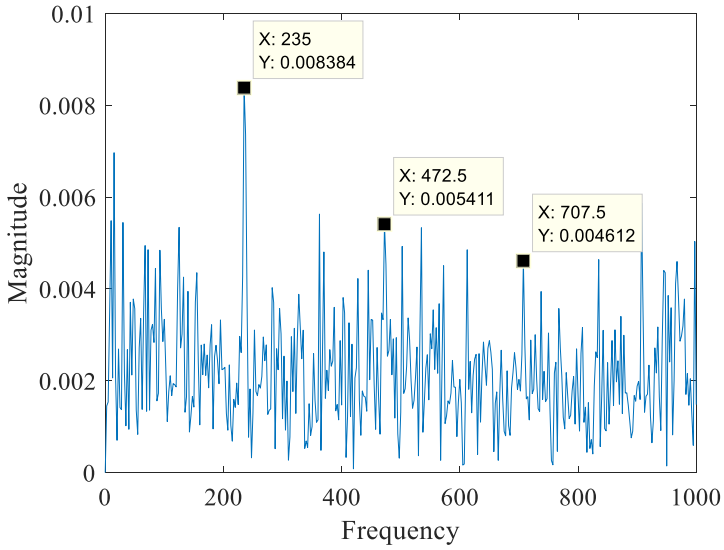


Figure 20 – Weak Fault Frequency by Wavelet

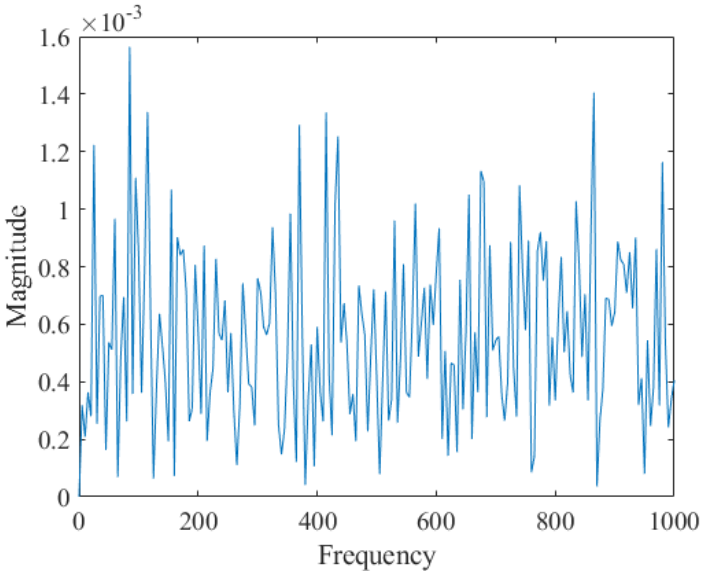


Figure 21 – Dual-Tree Discrete Wavelet Transform

The unscented Kalman filter was implemented to calculate the Q_1 and Q_2 values with the initial value of Q_1 and Q_2 selected to be 12 and 1.8. The goal of the Kalman filter is to maximize the kurtosis value generated by the TQWT with different Q_1 and Q_2. The resulted frequency plot is shown in Figure 22. It can be observed that the unscented Kalman

filter optimization algorithm yields similar resulted as the proposed method. However, the signal amplitude is attenuated.

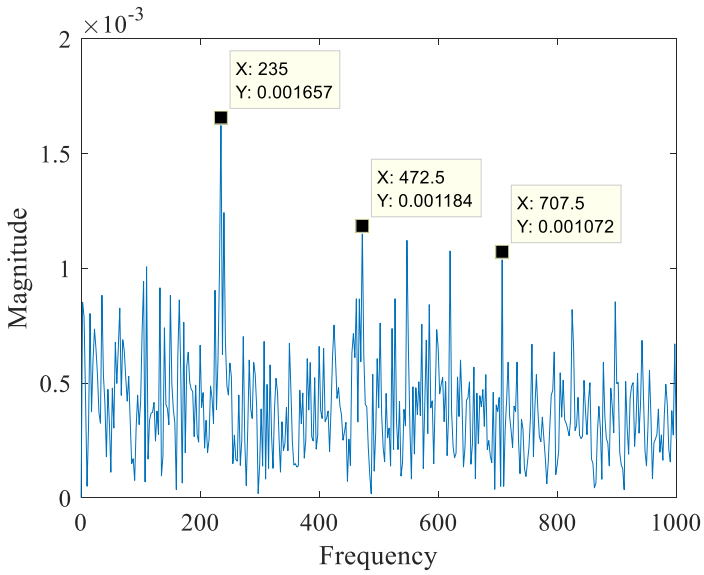


Figure 22 – TQWT Parameters Optimized by Unscented Kalman Filter

2.1.5 Conclusions

The detection of early fault is a challenging topic in the diagnosis of bearings. Documented method such as using kurtosis and root mean square value of the vibration signal could lead to inaccurate result due to the intrinsic stochastic nature of bearings. The integration of the DTWT+MCA model with the iterative alternating parameter selection can extract the desirable signal and its harmonics and improve the signal to noise ratio with respect to the traditionally used method. The proposed method is implemented to weak fault signal in the early degradation stage of the bearings. The DTWT achieves signal denoising using the complex wavelet transform similar to a band-pass filter and PCA helps select the optimal decomposed layer of the signal. The MCA splits the signal into High-Q and Low-Q parts. By tuning the parameters of the MCA, desirable fault frequency and its

harmonics can be extracted from the weak fault signal. In comparison, the traditionally used method cannot detect the weak fault signal as comprehensively as the proposed method in the early stage of the bearings. Simulated and experimental data are used to validate the proposed method. The proposed method achieves successful separations of the fault signal in terms of the signal to noise ratio in the validation data. In comparison with the widely used wavelet transform and dual-tree discrete wavelet transform, the dual-tree complex wavelet transform is able to extract fault signatures from signals with heavy noises. The proposed method has great significance in the field of diagnosis and prognosis because of its capability of the weak fault detection and flexibility of the parameter tuning process. Although the proposed method resolved the issue of parameter selection of the MCA, knowledge of the bearings' fault frequencies is still required. In addition, because of the nature of wavelet decomposition, the magnitude of the signal decreases with the increments of the decomposition layers. Critical information could be neglected based on the absolute magnitude of the signals. A potential improvement of this algorithm is to add different layers of the decomposed wavelet signal together to improve the magnitude of the fault signal and the signal to noise ratio. The effect of various wavelet basis can also be explored in the future studies. In addition, an unscented Kalman filter with possible constrains on the value of optimization parameters can be developed to further optimize the TQWT parameters.

2.2 DETECTION OF WEAK FAULT USING SPARSE EMPIRICAL WAVELET TRANSFORM FOR CYCLIC FAULT

2.2.1 Introduction

The initial stage of bearing diagnosis as proposed by Randall and Antoni [5] uses various techniques such as time-series model, adaptive line enhancer and time synchronous average to separate the discrete frequency noise. However, these methods have several disadvantages. The parameters of the time-series model is non-stationary for the vibrational signal of bearings as indicated in [42]. Therefore, an adaptive algorithm is required in general [43]. Time synchronous average normally requires tachometer or similar device to be installed on the equipment [1]. The cost and added complexity could make the order tracking and time synchronous average inapplicable. Although various research claims order tracking can be implemented only with vibration signal [44-47], the condition is very limited, and it is not applicable for weak fault.

McFadden implemented the high frequency resonance technique, also known as the envelope analysis, for fault detection of rolling element bearings [48]. This method extracts the bearing fault by examining the resonance generated by the impact of the fault site and is very effective in early fault detection [49]. Randall *et al.* established the relationship between the envelope analysis and the spectral correlation for cyclostationary machine signals [49]. This finding laid the foundation of the spectral analysis. Antoni proposed using the Spectral Kurtosis (SK) for analyzing non-stationary signals in [50]. The SK detects transients of vibrational signal regardless of presence of strong additive noise. Later, Antoni implemented the fast kurtogram in combination with envelope analysis for the extraction of fault frequency [51]. The model significantly reduces the computation of the traditional kurtogram [52] and improves the applicability for online monitoring.

The empirical mode decomposition (EMD) proposed by Huang *et al.* [53] is very effective for processing nonlinear and non-stationary signals in bearing diagnosis [54].

However, the lack of mathematical theory limits the method to be implemented in certain applications. Daubechies *et al.* proposed a model similar to the EMD to retrieve various modes of signal [55]. Although the EMD yields excellent results in processing nonlinear and non-stationary signals, the drawbacks of mode mixing, end effects and lack of criteria for sifting have created issues in the fault extraction of bearings. To address this problem, Torres *et al.* proposed the complete ensemble empirical mode decomposition method to alleviate the mode mixing of the traditional EMD and improve the computation efficiency [56]. The empirical wavelet transform (EWT) developed by Gilles combined the advantages of the EMD and wavelet transform [57]. The EWT overcame the deficiency of dyadic subdivision of the traditional wavelet by making the wavelet adaptive. It is shown to be effective in fault diagnosis of wind turbine [58].

In this section, a sparse based bearing diagnosis model is proposed. The model uses the kurtogram and envelope analysis to preprocess the signal to remove undesirable system noise and improve signal to noise ratio in the desired frequency range. A sparse empirical wavelet is implemented to obtain a sparse frequency-domain representation of the early weak fault signal and its second harmonic. Experimental data is used to validate the diagnostic model. The proposed method significantly enhanced the ability of the extraction of the weak fault signature by enhancing the signal to noise ratio in comparison with the traditional EWT.

2.2.2 Diagnostic Model

The proposed diagnostic model is shown in Figure 23. The vibration signal of bearing is first analyzed using the kurtogram to location the resonance frequency and

determine the bandwidth for the bandpass filter. A 100th-order AR filter is implemented to reduce the random noise in the system before performing the kurtogram. The segmentation of the signal in the frequency domain is obtained using a 1/3-binary tree of filter-banks to achieve better resolution on the frequency domain. The passband of the three additional quasi-analytic bandpass filter are $[0, \frac{1}{6}]$, $[\frac{1}{6}, \frac{1}{3}]$ and $[\frac{1}{3}, \frac{1}{2}]$ respectively. The resolution can be finely tuned by adding more bandpass filters with the increase of computation complexity. In this proposed model, a level of 7 is selected based on the 20480 Hz sampling frequency to achieve a minimum resolution of 26.67 Hz in the kurtogram. The signal is filtered based on the information obtained from the kurtogram, and the envelope analysis is implemented to locate the fault frequency and obtain the processed envelope signal. The envelope signal is then filtered by the sparse EWT to obtain a sparse solution of the frequency-domain signal which contains the fault signature.

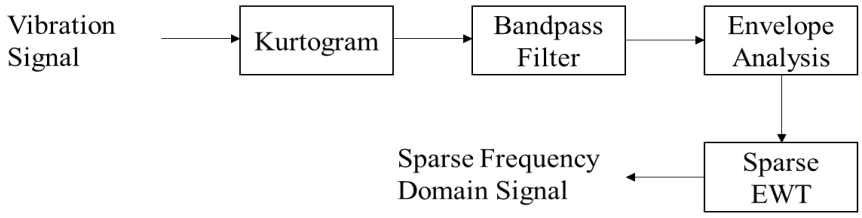


Figure 23 – Bearing Diagnostic Model

2.2.2.1 Kurtogram and Envelope Analysis

Spectral Kurtosis is suited for bearing diagnosis because of it capability to detect weak transient signal in additive noise. The implementation of the SK assumes that the vibration signal of bearings can be presented in (36):

$$y(t) = x(t) + n(t) \tag{36}$$

where $y(t)$ is the measured vibration signal, $x(t)$ is the fault signal with transients, and $n(t)$ is the stationary system noise. The SK is normally defined as (37) [52]:

$$SK_X(f) = \frac{\left\langle |H(n,f)|^4 \right\rangle}{\left\langle |H(n,f)|^2 \right\rangle^2} - 2 \tag{37}$$

where $H(n,f)$ is the complex envelope of signal at frequency *f*, and $\left\langle |H(n,f)|^4 \right\rangle$ is the temporary average of the envelope signal. The SK of a non-stationary signal is presented by (38) [52]:

$$SK_Y(f) = \frac{SK_X(f)}{[1 + SNR(f)]^2} \tag{38}$$

where $SK_Y(f)$ is the SK of the signal $y(t)$, $SK_X(f)$ is the SK of the signal $x(t)$, and $SNR(f)$ is the signal to noise ratio. When the signal to noise ratio is high, $SK_Y(f)$ is approximately equivalent to $SK_X(f)$. When the signal to noise ratio is low, $SK_Y(f)$ approaches zero. Therefore, by searching the whole frequency-domain, the SK is capable of distinguishing the fault signal from system noise.

The fast kurtogram examines a dyadic grid in the (*f, Δf*) plane instead of searching the whole plane to reduce the computation complexity. Before the fast kurtogram is implemented, the vibration signal is filtered by an auto-regressive filter of length of one

hundred to reduce transients. The first step of the fast kurtogram is to establish low-pass prototype filter $h(n)$ with cut-off frequency of $f_c = 1/8 + \varepsilon$, where $\varepsilon \geq 0$. The low-pass and high-pass filters from $h(n)$ are represented by,

$$h_l(n) = h(n)e^{j\pi n/4}, h_h(n) = h(n)e^{j3\pi n/4} \tag{39}$$

The two filtered are used to decompose the signal iteratively with each level consists of 2^k bands. The signal from the *i*th filter is presented by $c_k^i(n)$. The filtered signal by the low-pass and high-pass filter is denoted as $c_{k+1}^{2i}(n)$ and $c_{k+1}^{2i+1}(n)$, respectively at a decomposition level of *K-1*. The number of filtered signal is increased by a factor of 2 at each level. In addition, the respective length of signal is decreased by a factor of 2. The central frequency of the complex envelope of signal *x(n)* is represented by,

$$f_i = (i + 2^{-1})2^{-k-1} \tag{40}$$

and the bandwidth is denoted as,

$$(\Delta f)_k = 2^{-k-1} \tag{41}$$

The kurtosis is computed for all $c_k^i(n)$ for $i = 0,...,2^k - 1, k = 0,...,K-1$. Based on (37), the kurtogram equals to:

$$K_k^i = \frac{\left\langle \left| c_k^i(n) \right|^4 \right\rangle}{\left\langle \left| c_k^i(n) \right|^2 \right\rangle^2} - 2 \tag{42}$$

The envelope analysis has gained popularity in bearing diagnosis because of its effectiveness and simplicity of implementation [59]. The detailed methodology is presented in [49, 60]. The enveloped analysis in this section is carried on after the kurtogram by filtering the signal using a bandpass filter around the center frequency and squaring the filtered signal. The filtered signal is then low-pass filtered to obtain the envelope signal.

2.2.2.2 Sparse EWT

The empirical wavelet transform is implemented after obtaining the envelope signal from the kurtogram. The sparse EWT is composed of three parts: empirical mode decomposition, wavelet transform, sparse representation using the l_q norm. The empirical mode decomposition (EMD) first proposed by Huang *et al.* [53] has enabled extraction of features of bearing vibration signals [61]. The goal of the EMD is to decompose a signal $x(t)$ into different intrinsic mode functions (IMF) and residue *r* as shown in (43):

$$x(t) = \sum_{i=1}^{i=n} IMF_n + r \tag{43}$$

The IMFs contains various frequency bands corresponding to different frequencies presented in the system. The calculation of the IMFs is sensitive to the signal length. As the length of the signal increases, the computation effort increases significantly. Therefore, the adaptive boundary detection method is implemented to reduce computation effort. The initial division of the boundary is specified based on the frequency of interest, which is shaft frequency and outer race fault frequency and its harmonics. A significant number of

researches have demonstrated the advantage of using EMD for bearing diagnosis [62-64] and various derivations of the EMD method were proposed [56, 65]. The EWT has already demonstrated superiority in the estimation of mode and computation efficiency [66].

Wavelet transform is frequently used in signal processing of bearings to achieve the denoising of signal or extraction of features. The empirical wavelet uses the Littlewood-Paley and Meyer's wavelet shown in (44) and (45) [67]:

$$\psi(x) = (\pi x)^{-1}(\sin 2\pi x - \sin \pi x) \tag{44}$$

$$\psi_{m,n}^{b}(x) = 2^{-m/2}\psi(2^{-m}x - nb) \tag{45}$$

where m and n are an orthonormal basis for $L^2(\mathbb{R})$, and b is an arbitrary positive value. The assumption of the EWT separate [0, π] into N sections. ω_n is the limit between each section (where $\omega_0 = 0$ and $\omega_N = \pi$). The empirical scaling function and the empirical wavelet is denoted by (46) and (47) [57]:

$$\phi_n(\omega) = \begin{cases} 1, if\ |\omega| \le (1-\gamma)\omega_n \\ \cos\left[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega| - (1-\gamma)\omega_n)\right] \\ if\ (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0, otherwise \end{cases} \tag{46}$$

$$\psi_n(\omega)=\begin{cases} 1, if\ (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \cos\left[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_{n+1}}(|\omega|-(1-\gamma)\omega_{n+1})\right] \\ \quad if\ (1+\gamma)\omega_n \le |\omega| \le (1-\gamma)\omega_{n+1} \\ \sin\left[\frac{\pi}{2}\beta(\frac{1}{2\gamma\omega_n}(|\omega|-(1-\gamma)\omega_n)\right] \\ \quad if\ (1-\gamma)\omega_n \le |\omega| \le (1+\gamma)\omega_n \\ 0, otherwise \end{cases} \tag{47}$$

where $0<\gamma<1$ and $\beta(x)$ is an arbitrary function such that

$$\beta(x)=\begin{cases} 0, \ if \ \ x\le 0 \, and \, \beta(x)+\beta(1-x)=1 \ \ \forall x\in[0,1] \\ 1, \ if \ \ x\ge 1 \end{cases} \tag{48}$$

The $\beta(x)$ used in this model is

$$\beta(x)=x^4(35-84x+70x^2-20x^3) \tag{49}$$

The EWT is performed the same way as classic wavelet transform where the detail coefficients are obtained by calculating the inner products of signal with the empirical wavelets in (47). The approximation coefficients are obtained by calculating the inner products of signal with the scaling function. The reconstruction is known as (50):

$$f(t)=W_f^{\varepsilon}(0,t)\times\phi_1(t)+\sum_{n=1}^{N}W_f^{\varepsilon}(n,t)\times\psi_n(t) \tag{50}$$

where $W_f^{\varepsilon}(0,t)$ and $W_f^{\varepsilon}(n,t)$ are the approximation and detail coefficients, respectively. $\phi_1(t)$ is the scaling function, and $\psi_n(t)$ is the wavelet function.

Once the empirical wavelet transform is completed, the signal is reconstructed by the inverse empirical wavelet transform. Hilbert transform is performed to obtain the reconstructed signal in the frequency domain. A sparse representation of the frequency domain signal is obtained to reveal the fault information. The sparsity of the empirical wavelet transform is obtained by using the l_q regularized sparse regression on the frequency domain data with $0 \le q \le 1$. The l_q regularized sparse regression solve the optimization problem:

$$\min_x \|x\|_q \; subject\ to\ y = Ax \tag{51}$$

where $x \in \mathbb{R}^m$ is the reconstructed sparse signal, $y \in \mathbb{R}^m$ is the defect signal in the frequency domain and $A = 1/\sqrt{m} I_m$ with I_m as the $m \times m$ identity matrix and $\|x\|_q = \left(\sum_i |x_i|^q\right)^{1/q}$ with $0 \le q \le 1$. The l_q norm indicates that the reconstructed signal belong to an l_q ball for $0 \le q \le 1$, which yields a sparse reconstructed signal. This soft sparsity suggests that the *N* most important frequencies were extracted with the error $O(N^{\frac{1}{2}-\frac{1}{q}})$. In practice, we choose $q = 1$, which gives the Lasso, basis pursuit, or the Dantizig selector. A variety of algorithms have been developed for the l_q regularized sparse regression, such as the l_1 based method basis pursuit [68], the Lasso [69], and the

l_q based methods [70, 71]. The theoretical analysis of the reconstruction quality can be found in [72] as well as references therein.

2.2.3 Experiment

To validate the proposed method, the experimental data of a bearing run-to-failure test from Qiu *et al.* is used [16]. The second set from the testing data is used. The failure of the outer race is observed in bearing 1 at the end of the test, and theoretical failure frequency is 236.4 Hz. The overall vibration signal is shown in Figure 24. To detect the early fault signature, the signal around 4000 min shown in Figure 25 is used as weak fault signal to validate the proposed method.

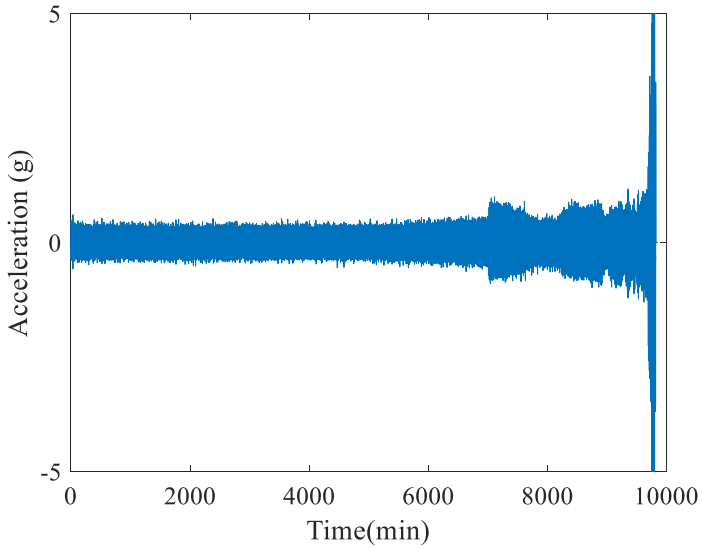


Figure 24 – Vibration Signal of Bearing Run-to-Failure Test

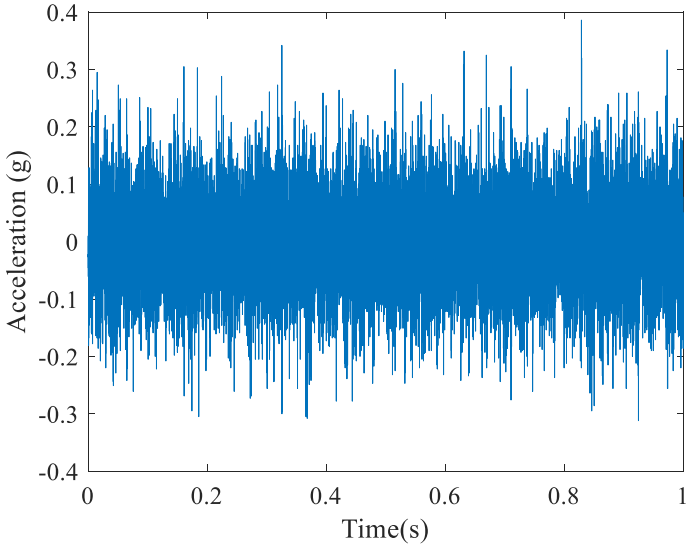


Figure 25 – Vibration Signal of Bearing around 4000 min

2.2.4 Result

The Hilbert transform of the unprocessed signal is shown in Figure 26. The fault signal is not distinctive in the unprocessed signal. The calculated kurtogram of the weak

fault signal is shown in Figure 27. The center frequency of the system resonance is determined to be 5880 Hz, and the bandwidth is calculated to be 80 Hz. The bandpass filter is set to keep signals within 5800 Hz and 5960 Hz. Two different levels of the signal are filtered. The levels are selected by inspecting the kurtogram and performing principal component analysis. The obtained signals are further filtered by a 16^{th} order Hanning window lowpass filter with the cutoff frequency equivalent to 40% of the Nyquist frequency. The filtered signal is squared to obtain the envelope signal.

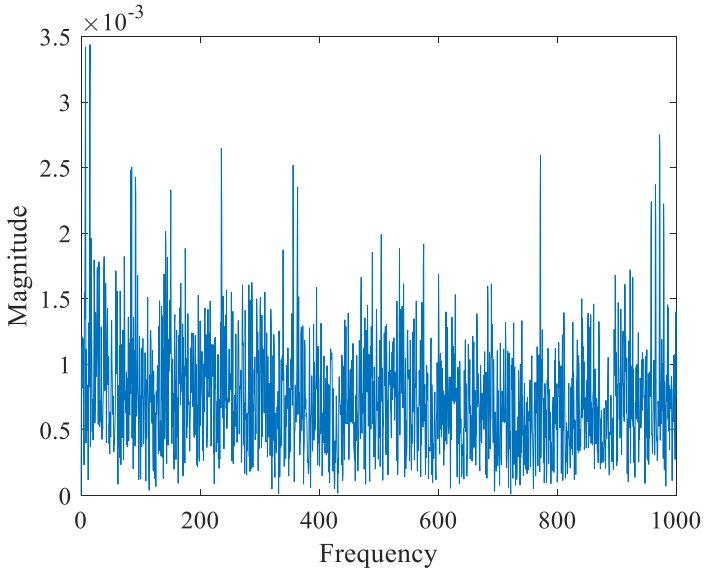


Figure 26 – Hilbert Transform of Unprocessed Signal

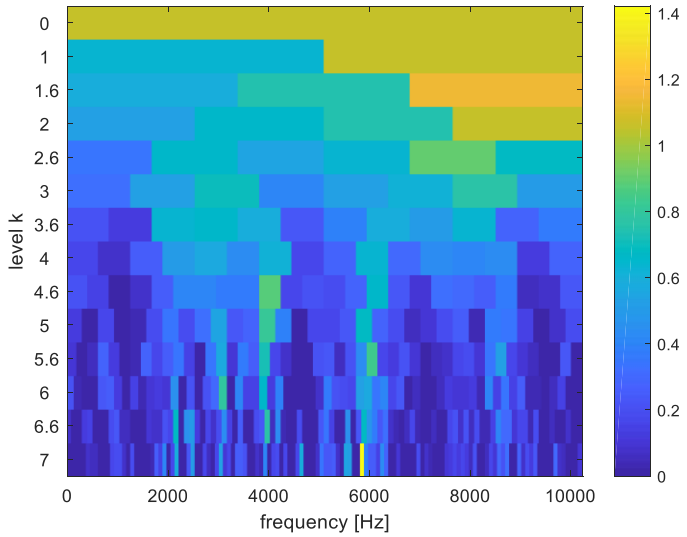


Figure 27 – Kurtogram of the Weak Fault Signal

The proposed sparse EWT is implemented to further enhance the signal to noise ratio of the defect signal. The extracted signal in the frequency-domain calculated using the Hilbert Transform is shown in Figure 28 and 29. The signal to noise ratio is defined as the amplitude of signal at defect frequency versus all other frequencies. The corresponding

signal to noise ratios of the two extracted frequencies are shown in Table 3 along with the signal to noise ratio of the unprocessed signal. The sparsity can be tuned by changing the degree of freedom of the l_q norm. In this case, a degree of freedom of 55 is chosen without sacrificing significant computation time and achieve desirable result. The traditional EWT is used as a benchmark to compare with the proposed method. It can be observed that the sparse EWT significantly reduces the random noises in the system and enhances the signal to noise ratio in Table 3.

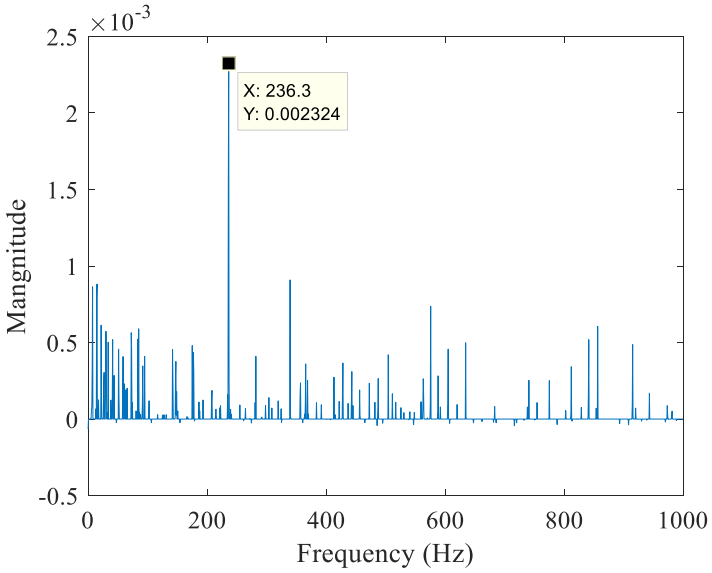


Figure 28 – Extracted Fault Frequency

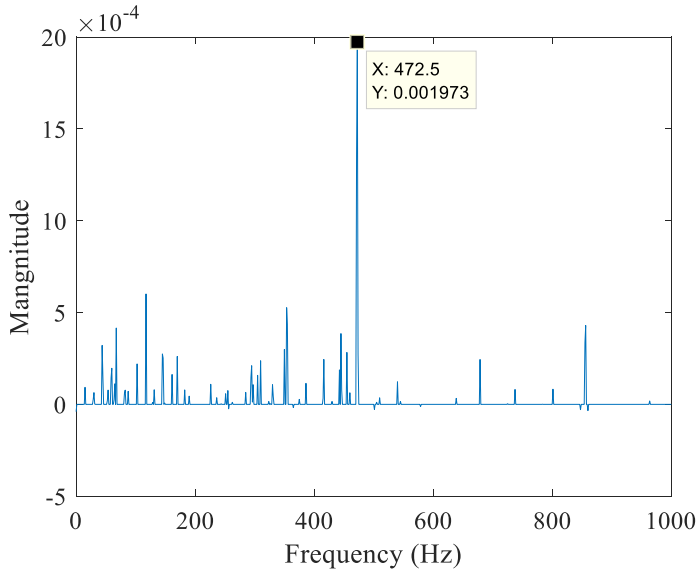


Figure 29 – Extracted Harmonic of Fault Frequency

Table 3 – Signal to Noise Ratio

Fault Frequency	Unprocessed Signal	Traditional EWT	Proposed Sparse EWT
236 ± 1 Hz	0.0054	0.0101	0.1339
472 ± 1 Hz	0.0015	0.0173	0.3396

2.2.5 Conclusion

This section proposed a bearing diagnostic model using the kurtogram and sparse EWT to extract single weak fault signature in rolling element bearings. The model uses the kurtogram and envelope analysis for preprocessing of the signal to eliminate the transient signals and locate the frequency band of the cyclic fault. The sparse EWT further filters out undesirable noises and improves the signal to noise ratio. The proposed model demonstrates a significant improvement of signal to noise ratio in comparison with the traditional EWT model. The model has a wide range of applications in detection of early cyclic fault. By proper tuning the parameters of the kurtogram, the model is not limited to detection of single fault and its harmonics. Although the sparse EWT model yields a satisfactory result in the single fault bearing diagnosis case, the model has a few drawbacks. The proposed model is very sensitive to the Fourier boundary detection method. Different methods could yield drastic changes in the detection accuracy and computation time. The effect of the wavelet parameter has not been investigated thoroughly, which are normally chosen by default or trial and error. An optimization algorithm could be added into optimizing the wavelet parameters. Future research can be conducted to add the automatic

examination feature of the kurtogram in order to implement this model in the online monitoring of the degradation process of various mechanical systems.

2.3 ADAPTIVE ONLINE DICTIONARY LEARNING FOR BEARING FAULT DIAGNOSIS

2.3.1 Introduction

In the last decade, the concept of condition-based maintenance (CBM) was proposed to better accommodate the industry requirements of machineries. A critical step to perform CBM is to accurately diagnose the severity and location of fault [6, 43, 73]. A general guideline for bearing diagnosis proposed by Randall and Antoni is shown in [5]. In the proposed tutorial, a phenomenon of strong masking signals from machine components is reported. Because of strong background noise generated by machine component and acquisition system, various signal processing techniques are implemented to reduce the system noise and extract the important information.

The most widely used methods are categorized into three different types: Time-domain, Frequency-domain and Time-Frequency analysis. The early research mainly focus on the Time-domain analysis. The most general method includes calculating the root mean square (RMS), skewness, kurtosis and peak values of the acquired vibration signal [7, 74-76]. The abovementioned methods still serve as the foundations of bearing diagnosis nowadays. Over the last few decades, the method of discrete/random separation is developed to remove the effect of small fluctuations in rotating speed for various machines. Bonnarot *et al.* implemented the unsupervised noise cancellation in the angular domain to resample the data in the time domain for the removal of random noise [77]. The method

significantly improves the signal to noise ratio of the vibration signal of bearings. McFadden implemented the time averaging method to remove the undesirable noise, but the proposed model is only limited to certain conditions [48]. Sawalhi and Randall implemented the autoregressive (AR) model to remove the regular gear meshing noise in [78]. The residual signal after the AR model reveals the periodical information of the rotating component within the gearbox. In general, the time-domain method has several disadvantages. The parameters of the time-series model various over the degradation process of rotating machineries as indicated in [42]. Therefore, an adaptive algorithm is required in general [79]. Although various time synchronous average methods are proposed, the method still relies on auxiliary devices to acquire information other than the vibrational signal [80]. Although several tacholess order tracking method can be implemented in limited conditions [13, 81, 82], The true tacholess order tracking method is still under development for early fault detection.

The frequency domain approach usually involves the Fourier Transform, Hilbert Transform and various demodulation techniques. The Hilbert Transform is the relationship between the real and imaginary parts of the Fourier Transform [5]. Rai and Mohant demonstrates the effectiveness of using Hilbert Transform incorporated with the empirical mode decomposition in bearing diagnosis [9]. The proposed method successfully extracted the fault signature while the traditional Fast Fourier Transform cannot achieve. Modulation occurs when a carrier signal's amplitude or frequency varies with respect to time [1]. Stack *et al.* implemented an amplitude modulation detector for the detection of bearing fault [83]. The method reduces its difficulty by only requiring the knowledge of bearings characteristic fault frequencies. Gong and Qiao implemented the current demodulation

technique to identify the bearing fault [84]. The proposed method has the advantage of ease of implementation because only the current need to be monitored.

The most widely used signal processing technique nowadays is the time-frequency domain analysis. The Short Time Fourier Transform, which is equivalent to performing Fourier Transform in a moving data window, is the simplest to implement [85]. Another popular technique is the wavelet analysis. Qiu *et al.* has successfully implemented the wavelet filter to detect weak signature in rolling element bearings [16]. Lei *et al.* combined the wavelet packet transform with Kurtogram to improve the performance of fault extraction [86]. Segreto *et al.* implemented the wavelet packet transform in combination with neural network to monitor vibration during the machining process [87]. A review of the applications of the wavelet transform can be found in [88]. A widely used technique for signal denoising nowadays is the Kurtogram. Kurtogram provides critical information for the spectral kurtosis analysis. Antoni first proposed using the Spectral Kurtosis (SK) to detect the transients of vibration signal in [50]. Later, Antoni implemented the fast Kurtogram in combination with envelope analysis to reduce the computation complexity while maintaining accurate result [51].

Online dictionary learning [89] is an efficient tool for signal denoising and feature extraction. The technique is first implemented to analyze multi-dimensional data [90]. It has the capability of adaptation to dynamic train sets and fast convergence. Over the years, various technologies are developed to train the dictionary and find the sparse representation of the signal to be analyzed. Aharon *et al.* presents the K-SVD method to calculate the dictionaries based on the training data and parameter [91]. Various versions and modifications to improve the computation speed of the K-SVD algorithms are presented in

[92-94]. Once the dictionaries are obtained, a sparse representation of the signal needed to be processed is calculated by the orthogonal matching pursuit (OMP) algorithm. A wide variety of the OMP algorithms are list in [95-97]. The dictionary learning methods have been used for bearing fault diagnosis [98, 99], which restricted diagnosis signal to shift-invariant in time and assumed hidden Markov model structure.

In this section, an adaptive online dictionary learning algorithm for bearing fault diagnosis is proposed. Comparing to the previous work using dictionary learning methods, our proposed work is free from assumptions on the shift-invariant property and allow online update for newly collected signals. The model uses the Kurtogram and envelope analysis to remove undesirable system noise and locate the fault frequency band. The dictionaries are first trained using a set of vibrational data with distinctive fault frequencies. After the training of the dictionaries, a new set of training data is input to the diagnostic model. A weighted least square algorithm is implemented to update the weights of the dictionaries rather than recalculating the dictionaries using the K-SVD algorithm. The diagnostic model is compared with a wavelet diagnostic model to demonstrate its capability of early fault detection.

2.3.2 Diagnostic Model

The proposed diagnostic model is shown in Figure 30. The vibration signal of the bearing is first denoised using the Kurtogram and envelope analysis [51]. The filtered signal is input into the online dictionary learning algorithm. The online dictionary learning algorithm calculates the desired dictionaries based on selected parameters. The dictionaries are tested using a series of data belong to the same training set to validate successful

training of the dictionary. The testing bearing vibration signal is used as the input for the OMP algorithm, and the OMP outputs the sparse representation of the testing signal. Once the dictionary training yield satisfactory result, the training for the dictionary is terminated. The dictionary update starts with a new set of training data. The sparse representation of the training data is obtained through the OMP algorithm. A weighted least square algorithm is implemented in the dictionary update process. The updated dictionary is able to distinguish additional fault frequency in the bearing vibration signal.

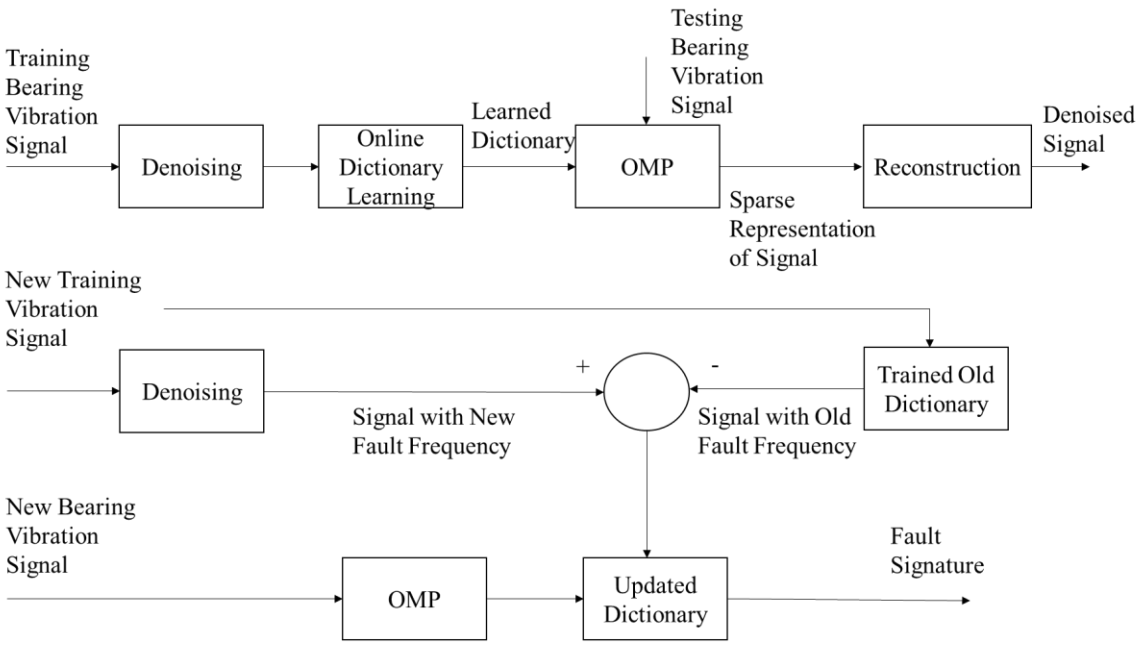


Figure 30 – Adaptive Bearing Diagnostic Model

2.3.2.1 Kurtogram

Spectral Kurtosis is suited for bearing diagnosis because of it capability to detect weak transient signal in additive noise. The implementation of the SK assumes that the vibration signal of bearings can be presented in (52):

$$y(t) = x(t) + n(t) \tag{52}$$

where $y(t)$ is the measured vibration signal, $x(t)$ is the fault signal with transients, and $n(t)$ is the stationary system noise. The SK is normally defined as (53) [52]:

$$SK_X(f) = \frac{\left\langle |H(n,f)|^4 \right\rangle}{\left\langle |H(n,f)|^2 \right\rangle^2} - 2 \tag{53}$$

where $H(n,f)$ is the complex envelope of signal at frequency *f*, and $\left\langle |H(n,f)|^4 \right\rangle$ is the temporary average of the envelope signal. The SK of a non-stationary signal is presented by (54) [52]:

$$SK_Y(f) = \frac{SK_X(f)}{[1 + SNR(f)]^2} \tag{54}$$

where $SK_Y(f)$ is the SK of the signal $y(t)$, $SK_X(f)$ is the SK of the signal $x(t)$, and $SNR(f)$ is the signal to noise ratio. When the signal to noise ratio is high, $SK_Y(f)$ is approximately equivalent to $SK_X(f)$. When the signal to noise ratio is low, $SK_Y(f)$ approaches zero. Therefore, by searching the whole frequency-domain, the SK is capable of distinguishing the fault signal from system noise.

The fast kurtogram examines a dyadic grid in the (*f, Δf*) plane instead of searching the whole plane to reduce the computation complexity. Before the fast kurtogram is implemented, the vibration signal is filtered by an auto-regressive filter of length of one

hundred to reduce transients. The first step of the fast kurtogram is to establish low-pass prototype filter $h(n)$ with cut-off frequency of $f_c = 1/8 + \varepsilon$, where $\varepsilon \geq 0$. The low-pass and high-pass filters from $h(n)$ are represented by,

$$h_l(n) = h(n)e^{j\pi n/4}, h_h(n) = h(n)e^{j3\pi n/4} \tag{55}$$

The two filtered are used to decompose the signal iteratively with each level consists of 2^k bands. The signal from the *i*th filter is presented by $c_k^i(n)$. The filtered signal by the low-pass and high-pass filter is denoted as $c_{k+1}^{2i}(n)$ and $c_{k+1}^{2i+1}(n)$, respectively at a decomposition level of *K-1*. The number of filtered signal is increased by a factor of 2 at each level. In addition, the respective length of signal is decreased by a factor of 2. The central frequency of the complex envelope of signal *x(n)* is represented by,

$$f_i = (i + 2^{-1})2^{-k-1} \tag{56}$$

and the bandwidth is denoted as,

$$(\Delta f)_k = 2^{-k-1} \tag{57}$$

The kurtosis is computed for all $c_k^i(n)$ for $i = 0,...,2^k - 1, k = 0,...,K-1$. Based on (53), the kurtogram equals to:

$$K_k^i = \frac{\left\langle \left| c_k^i(n) \right|^4 \right\rangle}{\left\langle \left| c_k^i(n) \right|^2 \right\rangle^2} - 2 \tag{58}$$

The envelope analysis has gained popularity in bearing diagnosis because of its effectiveness and simplicity of implementation [59]. The detailed methodology is presented in [49, 60]. The enveloped analysis in this section is carried on after the kurtogram by filtering the signal using a bandpass filter around the center frequency and squaring the filtered signal. The filtered signal is then low-pass filtered to obtain the envelope signal.

2.3.2.2 Online Dictionary Learning

The sparse representation of signals is a popular research topic in multi-dimensional signal processing and data compression in the recent years. By implementing an over-complete dictionary $D \in \mathbb{R}^{n \times k}$ which contains *n* number of rows and *k* number of columns of data, a signal $y \in \mathbb{R}^n$ can be represented as a linear combination of the dictionaries denoted as $y = Dx$, where $x \in \mathbb{R}^k$ is the sparse representation of the signal *y*. The object is to find either:

$$\min_x \|x\|_0 \quad subject\ to\ y = Dx \tag{59}$$

or

$$\min_x \|x\|_0 \quad subject\ to \|y - Dx\|_2 \le \varepsilon \tag{60}$$

where ε is the error tolerance, which is selected to be 0.01 to ensure relatively low computation effort while maintaining reconstruction accuracy. The online dictionary learning algorithm aims to find the sparser representations of the signal by solving the optimization problem defined in:

$$\min_{D,x} \|y - Dx\|_F^2 \quad subject\,to\, \forall i \|x_i\|_0 \le K \tag{61}$$

where K is the targeted sparsity [97].

The algorithm consists of two alternating steps: 1. Given current dictionary estimate, calculating the sparse representation matrix x which is normally done by using the OMP introduced in the next section. 2. Updating dictionary one atom at a time to optimize the target function defined in (61) for each atom individually. The resulting problem is a rank-1 approximation task shown in (62):

$$\{a,b\} := Arg\min_{a,b} \left\|E - ab^T\right\|_F^2 \quad subject\,to \|a\|_2 = 1 \tag{62}$$

where E is the error matrix, a is the updated atom and b^T is the new coefficients row in x. The problem can be directly solved by singular value decomposition (SVD) [100] or iterative updated method such as gradient descent or Newton methods [101].

2.3.2.3 OMP

The OMP algorithm implemented takes an $m \times n$ data matrix y, an $m \times l$ dictionary D and the desired sparsity level of k and outputs a $l \times n$ sparse representation of the data

matrix *y*. The estimate for the ideal signal is obtained by *Dy*. The general procedure usually involves:

1.Initialize the residual defined by $e = y - Dx$ as $e = y$, initialize a set Λ_p containing p elements from $\{1,...,l\}$ and set $\Lambda_0 = \varnothing$ and counter $t = 1$.

2. Find the index λ_t that solves the optimization problem defined in (63):

$$\lambda_t = Arg \max_{i=1,...,l} \left| e_{t-1}, \varphi_j \right| \tag{63}$$

where φ is the columns of *D*.

3. Augment Λ_p and the matrix of chosen atoms: $\Lambda_p = \Lambda_{p-1} \bigcup \{\lambda_t\}$ and $D_t = [D_{t-1} \varphi_{\lambda_t}]$

4. Solve a least square problem to get the new signal estimation:

$$y_t = Arg \min_y \left\| y - D_t y \right\|_2 \tag{64}$$

5. Calculate the new approximation of the data a_t and the new residual as:

$$a_t = D_t y_t \tag{65}$$

$$e_t = y - a_t \tag{66}$$

6. Increase *t*, and return to Step 2 until $t < p$.

7. The estimate *Dy* has nonzero indices at the components listed in Λ_p and the value of the estimation in component λ_i equals the *i*th component of y_t .

2.3.2.4 Dictionary Update

The dictionary weight update is performed using an iteratively reweighted least square algorithm [102]. The error E between the new signal Y and the reconstructed signal $D_{old}X$ using the old dictionary is shown in (67):

$$E = Y - D_{old}X \tag{67}$$

The weight is selected as shown in (68):

$$W = diag[\frac{1}{\|e_1\|_2^2},...,\frac{1}{\|e_n\|_2^2}] \tag{68}$$

where e_n is the individual term within *E*. The updated dictionary is:

$$D_{new} = YWX^T(XWX^T)^{-1} \tag{69}$$

The updating scheme is executed iteratively until the average value of the error *E* converges to the desirable value. The desirable value should be set with consideration of the actual applications. In this section, *E* is selected to be 0.50. By continuous updating the dictionary during the degradation process of bearings, more accurate diagnostic information can be obtained.

2.3.3 Experiment

To validate the proposed diagnostic method, the experimental data from Qiu *et al.* [16] is used. The bearings have a total of 32 rollers. The pitch diameter is 2.815 in., and the roller diameter is 0.331 in. The tapered contact angle is 15.2°. The shaft is driven by an AC motor at 2000 RPM. A 6000 lbs radial load is applied to the shaft and bearing through a spring mechanism as shown in the experimental setup. A magnetic plug is installed to collect the debris in the lubricant. Once the accumulated debris exceed the preset value, the test will stop. The data is taken every 10 minutes and the sampling frequency is 20480 Hz. The tested bearing had experienced outer race failure by the end of the test. The outer race failure frequency is calculated to be 236.4 Hz.

The overall vibration signal is shown in Figure 31. To simulate an online monitoring scenario, the bearing vibration data from 3900 min to 3990 min are used to train the dictionary. The training data is shown in Figure 32. To test the updating algorithm for the dictionary learning, the vibration data from 9750 min is used.

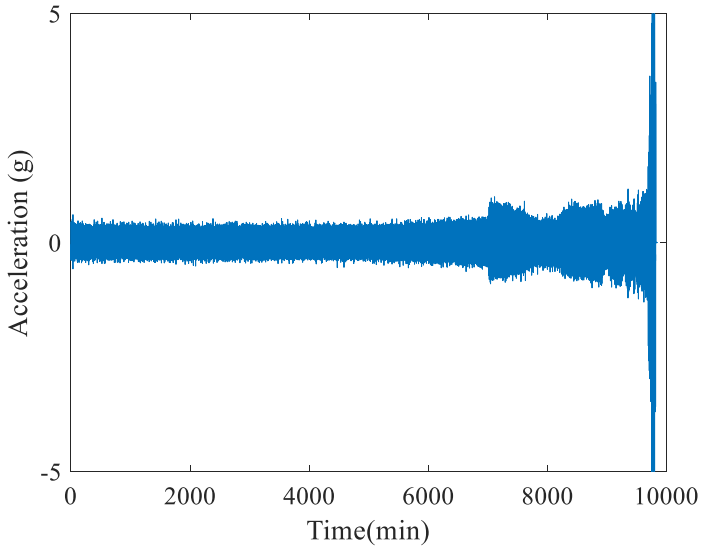


Figure 31 – Vibration Signal of Bearing Run-to-Failure Test

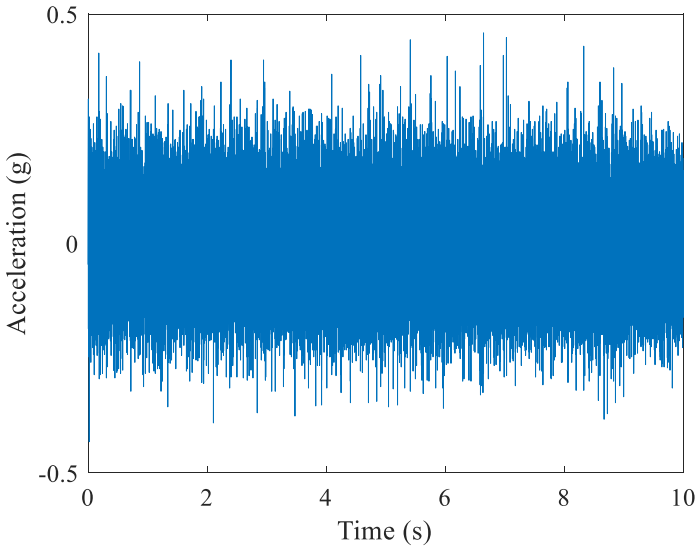


Figure 32 – Dictionary Training Data

2.3.4 Results

The training data are first analyzed by the Kurtogram. The mean is first subtracted from the vibration signal and a 16th order Hanning window lowpass filter with the cutoff

frequency equivalent to 40% of the Nyquist frequency based on the result from [50] is implemented to remove the random noise of the signal. The center frequency of the system resonance is determined to be 7680 Hz by the sampling rate and the decomposition level, and the bandwidth is calculated to be 5120 Hz [51]. The bandpass filter is set to keep signals within 2560 Hz and 12800 Hz. The Kurtogram for the training data is shown in Figure 33. Based on the magnitude of the Kurtogram, the filtered signal in level 1 is selected, and the amplitude spectrum of the filtered data is shown in Figure 34.

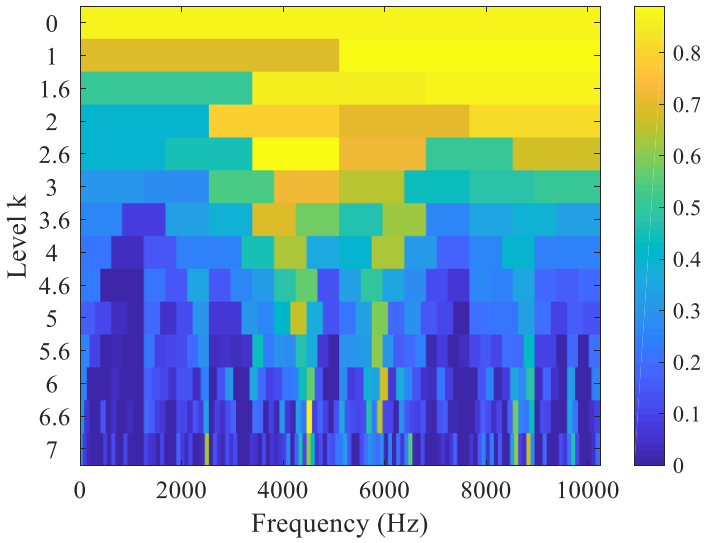


Figure 33 – Kurtogram of the Training Signal

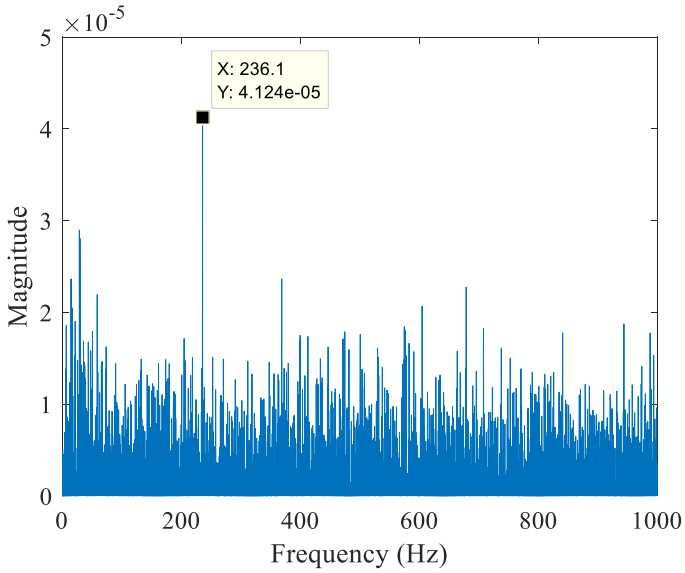


Figure 34 – Amplitude Spectrum of Filtered Data

It can be observed that by using the Kurtogram at the early degradation stage, the fault signature can be extracted. The filtered data is input into the K-SVD algorithm to train the initial dictionary parameter. The number of elements in each linear combination is set from 1 to 20. The number of dictionary element is set to be 50 with 15 iterations of

executing the K-SVD algorithm. The whole training is repeated for 20 times and for each time, the amplitude at the fault frequency and its 2nd harmonic is recorded. A criterion is established as:

$$Factor = \frac{A_{33\pm5} + A_{236\pm5} + A_{472\pm5}}{A_{all}} \tag{70}$$

where $A_{33\pm5}$ denotes the sum of amplitude of shaft rotating frequency interval, $A_{236\pm5}$ represents the sum of amplitude of the fault frequency interval, $A_{472\pm5}$ is the sum of amplitude of the harmonic frequency interval, and A_{all} is the sum of amplitude of all the frequencies together. The dictionary which yield the largest factor is selected for future signal denoising.

To demonstrate the effectiveness of the adaptive online dictionary algorithm, a new set of data with distinctive fault frequency, 2nd, 3rd and 4th harmonics are used in the dictionary learning algorithm. The filtered signal without updating the dictionary parameter is shown in Figure 35 with an average error of 0.59.

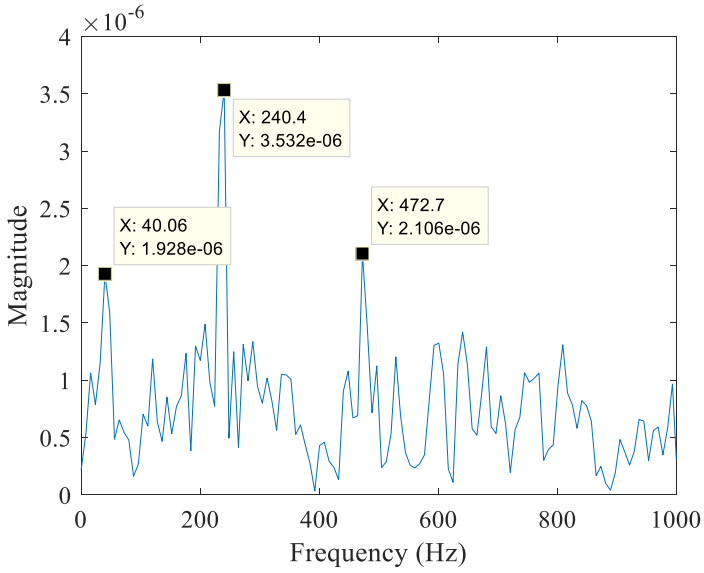


Figure 35 – Frequency Spectrum of Filtered Data Using Old Dictionaries

It can be observed that the dictionaries are still able to recover the significant information from the training set used before. However, it is not able to detect the added 3rd and 4th harmonics in the signal. After the dictionaries are updated using the proposed iteratively reweight least square method. The added harmonics can be detected using the new dictionaries as shown in Figure 36 with an average error of 0.4203

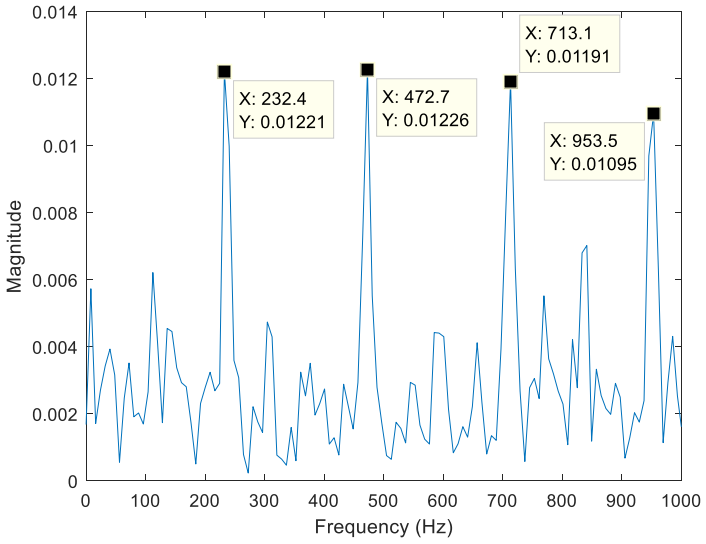


Figure 36 – Frequency Spectrum of Filtered Data Using New Dictionaries

By comparing Figure 35 and 36, it can be concluded that the update of the dictionary improves the detection of fault signature in frequency domain and increase the amplitude of the detected fault by a significant amount. In addition, more fault signatures are revealed by using this proposed dictionary updating method rather than recomputing the dictionary, which is time-consuming and cost inefficient.

2.3.5 Conclusions

In this section, an adaptive online dictionary learning bearing diagnostic model is proposed. The Kurtogram is initially implemented to find the weak fault signatures. The filtered signal is used to train the initial dictionaries using the K-SVD algorithm with parameters selected by trial and error. The trained dictionary is able to capture the features obtained by the Kurtogram without reperforming the analysis for a second time. As the damage of the bearing becomes more severe, new fault or harmonics are present in the system. The dictionary updating algorithm has the capability of distinguishing the newly

added fault feature whiles the old dictionary is not able to detect the newly developed damage. The adaptive online dictionary learning also has the benefits of short computation time because the dictionary training algorithm such as the K-SVD do not need to be performed with new training data. The proposed model creates a framework for the online monitoring of bearing degradation process. It is more robust and flexible than a fixed model such as the wavelet denoising, and it can combine the benefits of multiple denoising techniques. Future research will be performed on the detection of multiple fault signatures and vary the structures of the dictionaries.

2.4 BEARING FAULT DIAGNOSIS WITH NONLINEAR ADAPTIVE DICTIONARY LEARNING

2.4.1 Introduction

The continuous safe operation of the rotating machinery relies on the accurate knowledge of the machine health condition [103]. A significant amount of research effort has been invested into the monitoring of the health condition of rotating machinery in the recent years [5, 52, 75, 104]. The failure of rotating machinery is normally attributed to the bearing failure, the gear failure, rotating shaft fatigue, and the combination of these failure modes [7]. The inappropriate operation, severe operating condition, and inaccurate knowledge of the proper maintenance interval accounts for the major reasons that contribute to the catastrophic failure of the machines. The concept of condition-based maintenance or predictive maintenance has been proposed over a few decades [105]. It is still considered the most effective maintenance strategies until now as more effective diagnosis and prognosis techniques are developed [16, 53, 59, 106, 107]. The vibration

analysis of the vibration signal of bearings accounts for the majority of the current research conducted nowadays.

Because strong background noise signals generated by machine components and data transmission is observed in [108], the need of separation of the useful bearing vibration signal and noisy signal cannot be over emphasized. Various techniques are implemented to isolate the system noise and extract the critical information. The preprocessing are generally implemented using the tools of the following categories: Order tracking, time synchronous averaging, linear prediction, adaptive noise cancellation and deterministic/random signal separation [109]. The order tracking method generally requires information of the rotating shaft speed [110]. The vibration data is resampled based on the constant angular increments. The early documented research requires the angular increments to be relatively small to perform the order tracking [111]. The most recent research focuses on large angle variation and tacholess order tracking [13]. However, the technique requires complex computation algorithm to achieve a relative accurate result. The time synchronous averaging method, in general, requires the tachometer signal [112]. Although method without speed sensor exists [113], the no knowledge exists for effectively choosing angular resampling techniques. Trial and error methods are widely used, and the resampling techniques cannot be generalized [114]. Linear prediction is a relatively attractive way to obtain the deterministic part of the fault signal because of its simplicity [115]. The difficulty of implementation exists in the selection of an appropriate time series model and model order [116, 117]. The adaptive noise cancellation techniques are implemented to separate a signal containing two uncorrelated components such as gear fault and bearing fault [1]. It generally requires a

reference signal such as a known faulty gear vibration signal to successfully separate the mixture of signals. For the cases of system background noise and transmission error, the reference input of the model is difficult to determine. The discrete/random separation method does not require adaptation as the adaptive method. However, it requires the frequency components to be stable. In addition, the transfer function between signal and a delayed version of signal need to be computed, and the transfer function is sensitive to the signal to noise ratio [1]. In this section, the linear prediction method is implemented for the preprocessing of signal because of its simplicity and effectiveness for online monitoring applications.

The most widely implemented signal processing technique for the analysis of bearings' vibration is the time-frequency domain analysis. The Short Time Fourier Transform, which uses a moving data window to perform Fourier Transform, is the simplest to implement [1]. However, the resolution of the frequency domain signal is dependent on the selection of window size. The wavelet analysis is another popular and efficient way to perform time frequency analysis. A wide variety of the wavelet techniques exist such as the traditional wavelet transform [118], wavelet packet transform [119], complex wavelet transform [19]. Qiu *et al.* implemented the wavelet filter in the detection of weak fault signature of rolling element bearings [16]. The wavelet parameter is the key for early fault detection. However, no detailed guideline exists for the selection of the parameters. The most recent research focus on implemented intelligent algorithm to optimize the wavelet parameter [120]. A comprehensive review of the applications of the wavelet transform can be found in [88]. The Kurtogram [51] has become increasingly popular in the past few years for the analysis of bearing vibration signal because of its high

accuracy, ease of implementation, and wide range of applicability. The Kurtogram first uses filter banks to break down the signal into different frequency domains. Then, the kurtosis of the complex envelope of the filtered signal is calculated. The maximum kurtosis value is located, and the corresponding filtered signal is used for further analysis. A large value of the Kurtogram implies more impulsiveness in the signal to be analyzed. However, the Kurtogram has limitation of inability to reveal the cyclostationary characteristics of the signal [14]. This deficiency leads to the development of the Infogram which replaced the kurtosis with negentropy to reveal both the impulsive and cyclostationary characteristics of the vibration signal. This section implemented the Infogram as a part of the preprocessing tool to acquire the signal to update the dictionary.

The dictionary learning [89] is a powerful artificial intelligent tool for signal feature recognition and feature extraction. The technique is initially developed to analyze multi-dimensional data [90] for image recovery and denoising. The initiation of the dictionary learning requires calculations of the dictionary. Over the years, various technologies are developed to compute the dictionary and find the sparse representation of the signal to be analyzed [121-123]. Aharon *et al.* presents the k means singular value decomposition (K-SVD) method to compute the dictionaries based on the training data [122]. Later, various modifications have been made to the K-SVD algorithm to improve the computation efficiency [92-94]. After the computation of the dictionaries, a sparse representation of the target signal needed to be obtained. The sparsity of the signal is achieved by the implementation of the orthogonal matching pursuit (OMP) algorithm [95]. Similarly to the K-SVD, the OMP algorithm has a wide variety of variations listed in [95-97, 124] to ease computation effort. As mentioned in [122] a possible way to improve the training of the

dictionary is assigning weights to the different dictionary. In this case, the weights of the dictionaries can be adjusted by adaptive algorithm and possibly achieve structural adaptation of the dictionaries by eliminating the dictionaries with small weights.

In this section, a nonlinear adaptive dictionary learning algorithm for bearing signal fault diagnosis is proposed. Instead of using the traditional K-SVD method to train the dictionary for multiple times, the Unscented Kalman Filter (UKF) is implemented to achieve the adaptation of the dictionary. A nonlinear relationship between the dictionary and the artificial kurtosis is established to achieve the self-adaptation. Compared to the previous documented research, the proposed work does not require the learning of the dictionaries, is free from assumptions on the shift-invariant property, and allows online update for newly collected signals. The model utilizes the AR linear predictor to achieve random signal separation. The Infogram and envelope analysis is implemented to locate the fault frequency band and acquire the training data for the dictionaries. The dictionaries are initialized using a set of random variables and updated using the UKF. A sparse representation of the frequency domain signal is obtained in the end to reveal the fault signatures.

2.4.2 Diagnostic Model

The proposed adaptive bearing diagnostic model is shown in Figure 37. The training vibration signal of the bearing is preprocessed by the AR linear predictor to reduce the system noise. The AR model order is selected by trial and error. The autocorrelation function and partial autocorrelation function are obtained to facilitate the AR order selection. The wavelet packet transform is implemented to dyadically decomposed the

filtered signal. The square envelope (SE) and square envelope spectrum (SES) of the signal is obtained, and the negentropy of the signal is calculated. The band-passed signal with the maximum negentropy is selected as the training data for the dictionary. The UKF takes the weights of the dictionaries and the artificially created kurtosis as input, and calculates the updated weights of the dictionary. The updated dictionaries are tested using a set of vibration signal with weak fault to validate the diagnostic model.

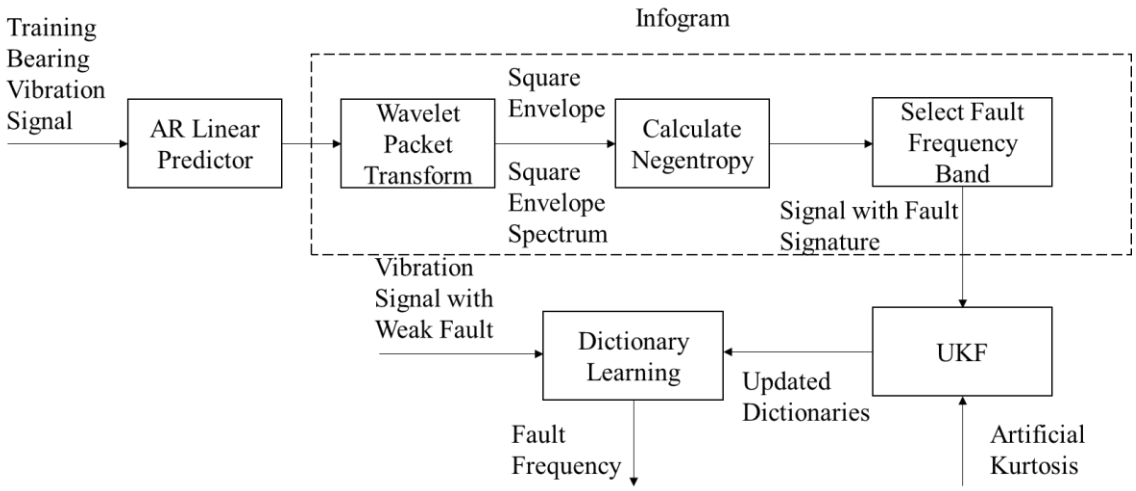


Figure 37 – Adaptive Bearing Diagnostic Model

2.4.2.1 Infogram

The Infogram is very similar concept to the Kurtogram with the introduction of the concept of entropy borrowed from thermodynamics. It aims to compensate for the lack of detectability of cyclostationary signal of the Kurtogram. The signal transient is perceived as a deviation from the equilibrium state of the squared envelope, and the cyclic characteristics introduces perturbation in the squared envelope spectrum. Based on observation from thermodynamics, the concept of negentropy is proposed by Antoni [14] to locate the frequency band that contains the most useful information. The implementation

of the Infogram is based on the assumption that the vibration signal of bearings can be separated as shown in (71):

$$y(t) = x(t) + n(t) \tag{71}$$

where $y(t)$ is the measured vibration signal, $x(t)$ is the fault signal with impulsiveness and cyclostationarity, and $n(t)$ is the system noise. The instantaneous flow of energy in the selected frequency band $[f - \Delta f / 2; f + \Delta f / 2]$ is represented by the SE as (72) [14]:

$$e_x(n; f, \Delta f) = \left| x(n; f, \Delta f) \right|^2 \tag{72}$$

where $x(n; f, \Delta f)$ is the vibration signal of the bearing at the corresponding frequency band, n is the length of the signal, and f is the center frequency of the bandpass filter. The average value of the energy fluctuation is represented by (73) [14]:

$$\bar{e}_x(f, \Delta f) = \frac{1}{L} \sum_{1}^{L} e_x(n; f, \Delta f) \tag{73}$$

where L is the length of the signal after the bandpass filterbank created by wavelet packet transform. The strength of the energy fluctuation can be represented by the variance of the energy flow which is (74) [14]:

$$Var_x(f; \Delta f) = \frac{\left\langle e_x(n; f, \Delta f)^2 \right\rangle}{\bar{e}_x(f, \Delta f)^2} - 1 \tag{74}$$

When the system has reached equilibrium, the value of the variance equals to 1, and a non-stationary signal will result in a variance value larger than 1. The connection of non-stationarity and nonlinearity for a zero-mean signal can be related to the spectral kurtosis shown in (75):

$$K_x(f;\Delta f) = Var_x(f;\Delta f) - 1 = \frac{\left\langle \left| x(n;f,\Delta f) \right|^4 \right\rangle}{\left\langle \left| x(n;f,\Delta f) \right|^2 \right\rangle^2} - 2 \tag{75}$$

The spectral kurtosis is reported to be able to account for a single or a few transients in [14], however, the impact generated by the bearing defect results in multiple repetition of the transient. The discrete Fourier transform of the square envelope signal, also known as the square envelope spectrum, is shown to be effective in transients detection [125]. The square envelope spectrum is denoted as (76) for discrete signal:

$$E_x(a;f,\Delta f) = \sum_{n=0}^{L} e_x(n;f,\Delta f)\exp(\frac{-j2\pi\alpha n}{F_s}) \tag{76}$$

where α is the frequency, F_s is the sampling frequency. With the addition of the square envelope spectrum, the Infogram is capable of processing impulsive and cyclostationary signal at the same time.

The negentropy, in replacement of the kurtosis, is calculated as (77) and (78) for the SE and SES signal [126]:

$$I_e(f,\Delta f)=\sum_{n=1}^{L}\frac{e_x(n;f,\Delta f)^2}{\frac{1}{N}\sum_{n}^{L}e_x(n;f,\Delta f)^2}\ln\frac{e_x(n;f,\Delta f)^2}{\frac{1}{N}\sum_{n}^{L}e_x(n;f,\Delta f)^2} \tag{77}$$

$$I_E(f,\Delta f)=\sum_{n=1}^{L}\frac{\left|E_x(\alpha;f,\Delta f)\right|^2}{\frac{1}{N}\sum_{n}^{L}\left|E_x(\alpha;f,\Delta f)\right|^2}\ln\frac{\left|E_x(\alpha;f,\Delta f)\right|^2}{\frac{1}{N}\sum_{n}^{L}\left|E_x(\alpha;f,\Delta f)\right|^2} \tag{78}$$

The frequency band with the maximum negentropy is selected as the training data for the dictionary learning.

2.4.2.2 Dictionary Learning

Dictionary learning as a recent popular tool decomposes a multi-dimensional signal as a linear combination of k interpretable vectors and compresses the signal in a sparse representation. The collection of such k vectors is referred to a dictionary, and each vector ina dictionary is called an atom. The dictionary learning considers a series of n-dimensional input signal $y\in\mathbb{R}^n$ and optimize a certain loss function to get an over-complete dictionary $D\in\mathbb{R}^{n\times k}$ and the corresponding coefficients $x\in\mathbb{R}^k$. By introducing a sparse regularization term in the loss function, we put a sparsity constraint on the coefficient so that the $x\in\mathbb{R}^k$ is a sparse representation of the input signal. The sparsest representation is to apply l_0 norm regularization in the loss function,

$$L(x,D)=\frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\left\|y_i-Dx_i\right\|^2+\rho\left\|x_i\right\|_0\right) \tag{79}$$

where m is the number of input signals with $i = 1,...,m$ and the regularization parameter ρ is a trade-off between the sparsity of x_i and the goodness-of-fit, which is specified by the user. D represent the dictionaries to recover the signal. Solving the l_0 norm regularization is known to be NP hard, which is not solvable for high dimensional case, i.e. k is large. As others, the relaxation approaches that use l_1 norm regularization in the loss function are also widely adapted:

$$L(x, D) = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{1}{2}\|y_i - Dx_i\|^2 + \rho\|x_i\|_1\right) \tag{80}$$

which is also known as *Lasso* or *basis pursuit*. The regularization parameter is specified to a value which requires relatively low sparsity while preserving the accuracy of the signal reconstruction. The dictionary learning algorithms aim at finding the sparse representation coefficients and the dictionary alternatively by solving the optimization problem defined in (81) or equivalently:

$$\min_{D,x}\|y - Dx\|^2 \; subject\, to \|x_i\|_0 \le d, \forall i = 1,...,m \tag{81}$$

where d is the targeted sparsity that is equivalent to ρ [97]. More specifically, the algorithm consists of two alternating steps: 1. Given current estimation of dictionary, finding the sparse representation coefficient x by solving the regularized least squares problem using method such as the OMP introduced in the next section. 2. Updating dictionary atom-wise to optimize the target function defined in (81) for each atom individually, which results in a rank-one approximation problem:

$$\{a,b\} := Arg\min_{a,b} \left\| E - ab^T \right\|_F^2 \quad subject\,to \left\| a \right\|_2 = 1 \tag{82}$$

where E is the error between input signal matrix and current recovered signal matrix, a is the updated atom and b^T is the corresponding row of new coefficients in x. The rank-one approximation problem can be either solved by iterative updated method such as gradient descent or Newton methods [101] or by direct solver such as singular value decomposition (SVD) [100].

2.4.2.3 Orthogonal Matching Pursuit (OMP)

Given an $m \times n$ data matrix y and the $m \times k$ dictionary matrix D as input, the OMP algorithm computes $k \times n$ sparse representation coefficient matrix x with desired sparsity level of d. The estimate for the recovered signal is Dx, and the residuals are defined by $e = y - Dx$. As an iterative greedy algorithm, the OMP selects at each step the atom that is most correlated with the current residuals. The OMP updates the residuals by projecting the signal y onto the space spanned by the selected atoms. In particular, the OMP for signal recovery can be summarized as following:

1. Initialize the residual as $e_0 = y$, an active set Λ_p containing p elements from $\{1,...,k\}$, and set $\Lambda_0 = \varnothing$ and counter for iteration as $t = 1$.

2. Find the atom φ_{j_t} that solves the optimization problem:

$$\varphi_{j_t} = Arg \max_{j \in \{1,...,k\}} \left| \varphi_j^T e_{t-1} \right| \tag{83}$$

and add the atom φ_{j_t} to the set of selected atoms in D_t.

3. Update $\Lambda_p = \Lambda_{p-1} \bigcup \{j_t\}$ and $D_t = [D_{t-1} \varphi_{j_t}]$.

4. Update the signal estimation through the least square optimization:

$$y_t = Arg \min_y \left\| y - D_t y \right\|_2 \tag{84}$$

5. Calculate the new approximation of the data a_t and the new residual as:

$$a_t = D_t y_t \tag{85}$$

$$e_t = y - a_t \tag{86}$$

which is the projection onto the linear space spanned by the active atoms.

6. Set $t = t + 1$, and return to Step 2 until $t < p$.

Note that the estimate *Dy* has nonzero indices at the active set Λ_p , where the estimation in component j_t equals the *i*th component of y_t .

2.4.2.4 Dictionary Update using UKF

With observed signal *y* and sparse representation coefficient *x* obtained by OMP, we propose an innovative dictionary update algorithm, based on artificially created kurtosis, which enables online application without human intervene. More specifically, instead of calculating dictionary as shown in (82) using computationally intensive

algorithm, the dictionary is updated using the UKF proposed by Julier and Uhlman [127]. The UKF uses a 3rd order polynomial to account for the nonlinearity versus the traditional extended Kalman filter (EKF) which uses 1st order approximation. The basic framework for the UKF is similar to the EKF which consists of estimation of the state of a nonlinear system [128]:

$$\theta_{k+1} = F(\theta_k, v_k) \tag{87}$$

$$y_k = H(\theta_k, n_k) \tag{88}$$

where θ_k denotes the intermediate unobserved state of the system, y_k is the observed signal, v_k is the process noise which drives the dynamic system, and n_k is the observation noise. The parameter estimation is achieved by assume a nonlinear mapping as (89):

$$y_k = G(\theta_k, w) \tag{89}$$

where θ_k is the input, y_k is the output, and the nonlinear function G is parameterized by the weight w. The error e_k is defined by (90):

$$e_k = d_k - G(\theta_k, w) \tag{90}$$

where d_k is the desired output. The goal is to solve parameter w which minimize the square error.

The unscented transformation assumes θ (dimension L) has a mean value of $\bar{\theta}$ and covariance $\boldsymbol{P}_x$. To obtain the statistics of the observed signal y, a matrix of $2L+1$ sigma vector χ_i is formed according to the following conditions:

$$\chi_0 = \bar{\theta} \tag{91}$$

$$\chi_i = \bar{\theta} + \left(\sqrt{(L+\lambda)\boldsymbol{P}_x}\right)_i \quad for\ i = 1,\ldots,L \tag{92}$$

$$\chi_i = \bar{\theta} - \left(\sqrt{(L+\lambda)\boldsymbol{P}_x}\right)_{i-L} \quad for\ i = L+1,\ldots,2L \tag{93}$$

$$W_0^m = \frac{\lambda}{L+\lambda} \tag{94}$$

$$W_0^c = \frac{\lambda}{L+\lambda} + (1-\alpha^2+\beta) \tag{95}$$

$$W_i^m = W_i^c = \frac{1}{2(L+\lambda)} \quad for\ i = 1,\ldots,2L \tag{96}$$

where the scaling factor λ is defined in (97):

$$\lambda = \alpha^2(L+\kappa) - L \tag{97}$$

in which α determines the spread of the sigma points around the mean $\bar{\theta}$. It usually has a relatively small value such as 0.001. κ is known as a secondary scaling parameter which

is normally 0, and β is used to embed prior information of the distribution of θ. The sigma vectors are propagated through the nonlinear function defined as:

$$Y_i = g(X_i) \quad for\ i = 0,...,2L \tag{98}$$

The mean and covariance for *y* are estimated using a weighted sample mean and covariance of the posterior sigma points as:

$$\bar{\boldsymbol{y}} \approx \sum_{i=0}^{2L} W_i^m Y_i \tag{99}$$

$$\boldsymbol{P}_y \approx \sum_{i=0}^{2L} W_i^c \left(Y_i - \bar{\boldsymbol{y}}\right)\left(Y_i - \bar{\boldsymbol{y}}\right)^T \tag{100}$$

The UKF is an extension of the UT to recursive estimate:

$$\theta_k = (prediction\ of\ \theta_k) + \kappa_k \cdot [y_k - (prediction\ of\ y_k)] \tag{101}$$

The *i*th dictionary is updated according to (102):

$$D_{i+1} = D_i \times \theta_k \tag{102}$$

In this section, the artifical kurtosis is created using a linear function. Random weights generated by a normal random number generator with mean value of 0 and standard deviation of 1 are assigned to each dictionary. The artifical kurtosis is treated as observation data, and θ_k is assigned as the actual state similar to [129].

2.4.3 Simulation and Experiment Result

To validate the proposed diagnostic method, the Infogram is first tested by a set of simulated signal from [120]. The AR filter length is set as 30 in this case. The resonance frequency is 3000 Hz, and the fault frequency is 100 Hz. The sampling frequency of the signal set to be 12000 Hz. The time domain representation of the simulated signal is shown in Figure 38, and the AR filtered signal is shown in Figure 39. The Hilbert transform of the filtered simulated signal is shown in Figure 40. It can be observed from Figure 39 that the Hilbert transform does not indicate

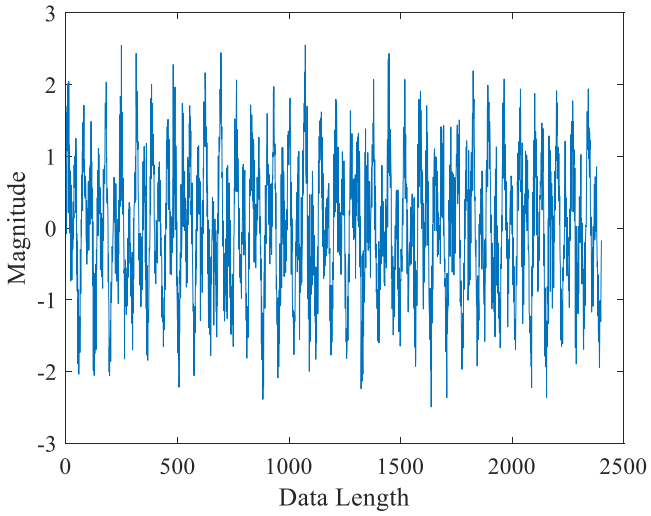


Figure 38 – Simulated Signal of Bearing Vibration

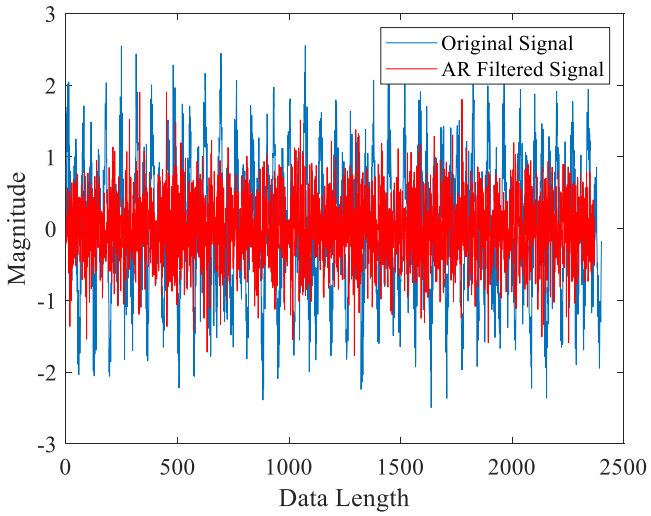


Figure 39 – AR Filtered Signal

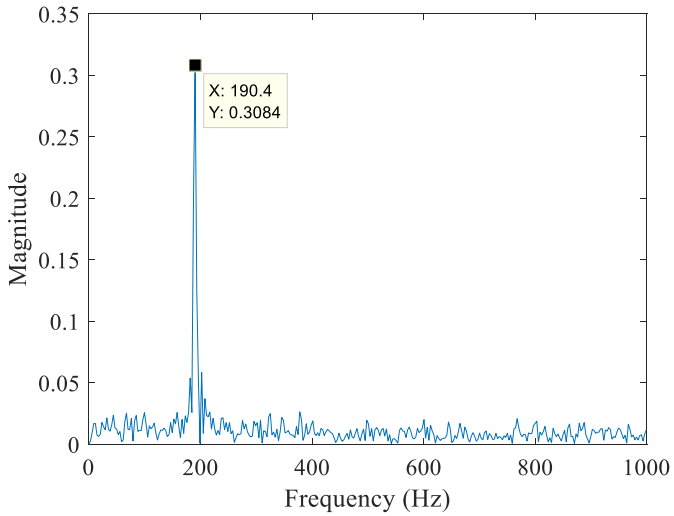


Figure 40 – Hilbert Transform of the Simulated Signal

the right fault frequency. The Infogram in (77) and (78) is performed on the simulated data and the SE and SES negentropy plots are shown in Figure 41 and 42. The center frequency for Figure 41 is calculated to be 3000 Hz which correspond to the resonance frequency,

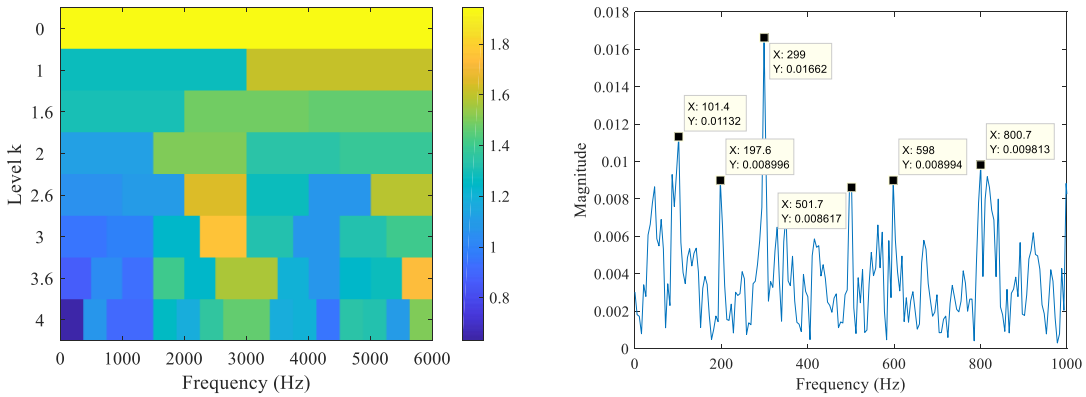


Figure 41– (a) Infogram for SE (b) Frequency Domain of the Filtered Signal by the SE Infogram

and the fault frequency and its harmonics are revealed in Figure 41 (b) in comparison to the one in Figure 40.

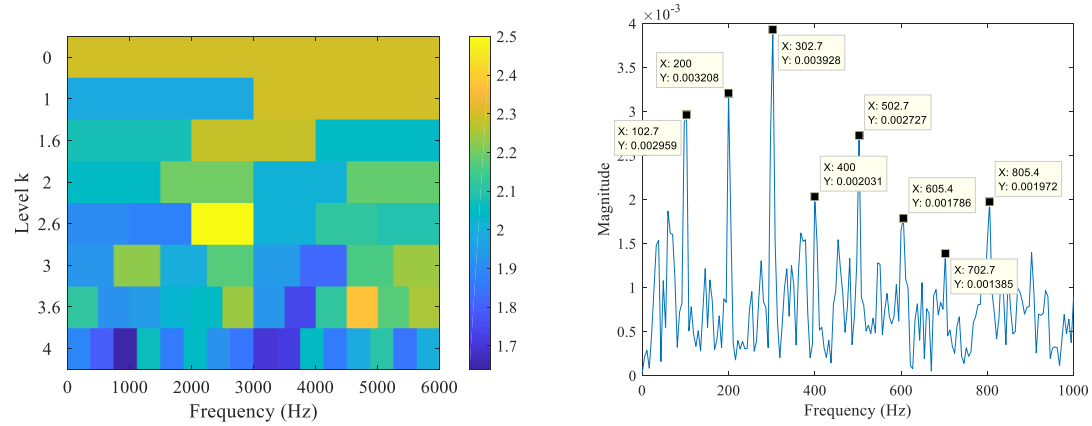


Figure 42 – (a) Infogram for SES (b) Frequency Domain of the Filtered Signal by the SES Infogram

The SES Infogram shown in Figure 42 has a center frequency of 2500 Hz, and the maximum negentropy value occurs at level of 2.5. The corresponding frequency plot also indicates the fault frequency of 100 Hz and its harmonics.

The experimental data from Qiu *et al.* [16] is used to validate the Infogram and the nonlinear adaptive dictionary learning. The training vibration signal acquired in between 5000 min and 5010 min as shown in Figure 43 is used for the dictionary training.

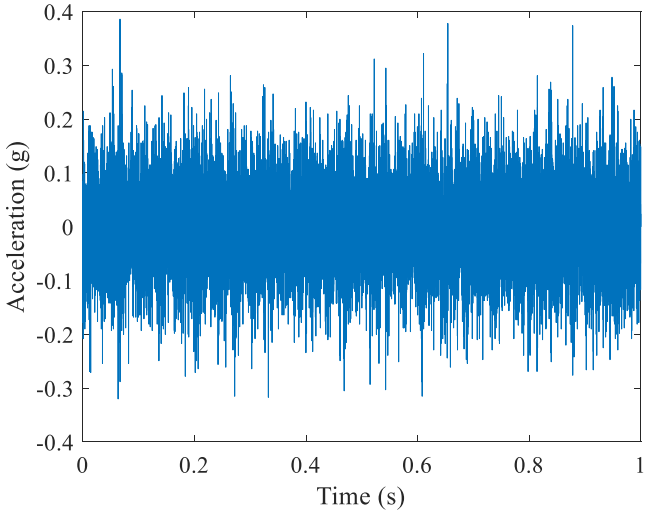


Figure 43 – Vibration Signal of Bearing Run-to-Failure Test

To properly determine the order of the AR filter, the autocorrelation and partial autocorrelation [130] is computed for lag value up to 500 as shown in Figure 44 and 45.

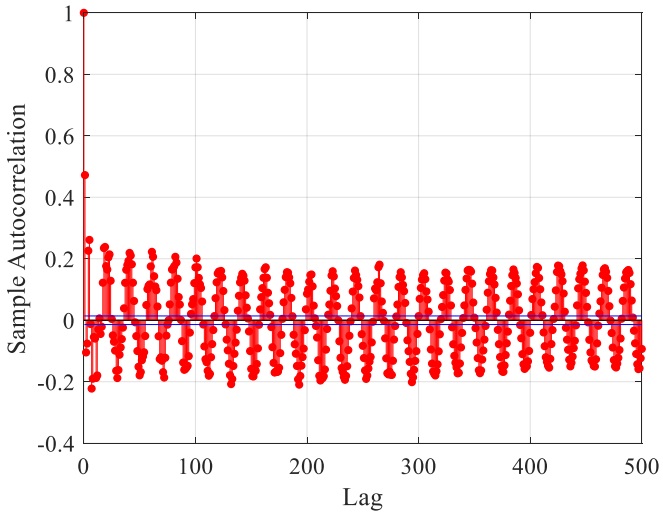


Figure 44 – Autocorrelation of Vibration Signal

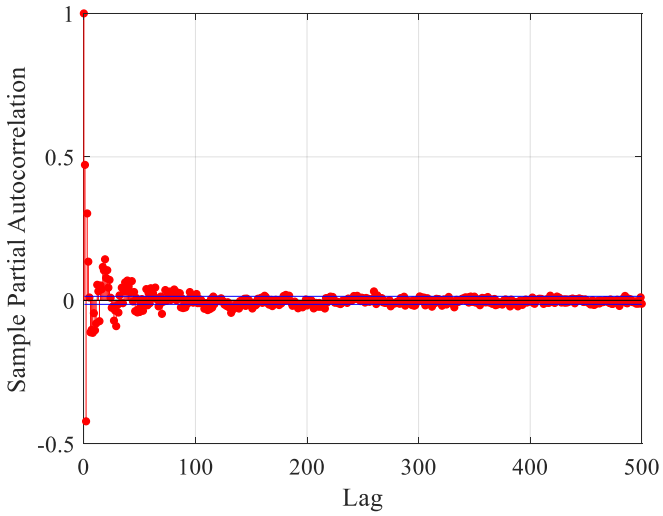


Figure 45 – Partial Autocorrelation of Vibration Signal

Based on the information obtained from the autocorrelation and partial autocorrelation, the proper estimation model should be anti-persistent. However, the anti-persistent series generally need the fractional ARIMA model, which significantly increases the computation complexity, to have relatively good prediction result. To reduce the computation time and simplify the linear prediction model, the AIC and BIC criteria is

implemented instead of the autocorrelation function for multiple AR, ARMA and ARIMA models to select the best model order. The AR model with an order of 30 is tested to have the lowest AIC and BIC value, which indicates the best fit. The filtered signal is shown in Figure 46.

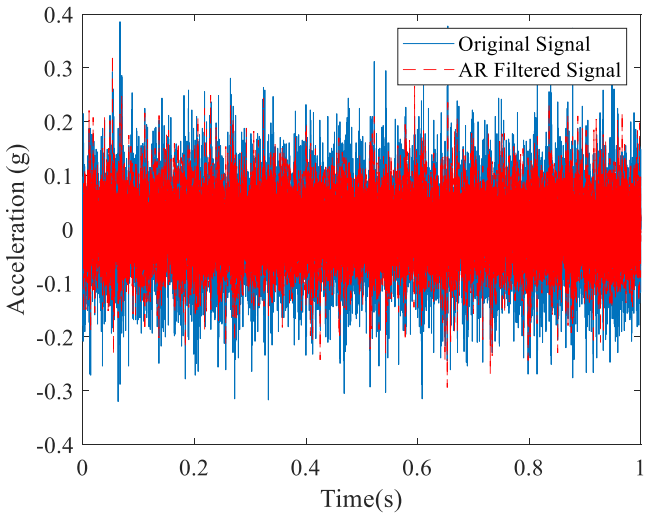


Figure 46 – AR Filtered Vibration Signal

The signal is analyzed by the Infogram using the wavelet packet transform as the filter bank and the SE and SES Infogram are shown in Figure 47 and 48 respectively. The SE Infogram indicates a center frequency of 5120 Hz, and the maximum negentropy value of 2.3 occurs at the decomposition level 0. The fault frequency and the shaft harmonic frequency are detected as shown in Figure 47 (b). The SES Infogram indicates a center frequency of 8533 Hz, and the maximum negentropy value of 2.4 occurs at decomposition level 1.5. In addition to the impulsiveness, the SES Infogram captures the second harmonics of the fault signal which agrees with the theory that the cyclostationary characteristics are dominated in the SES.

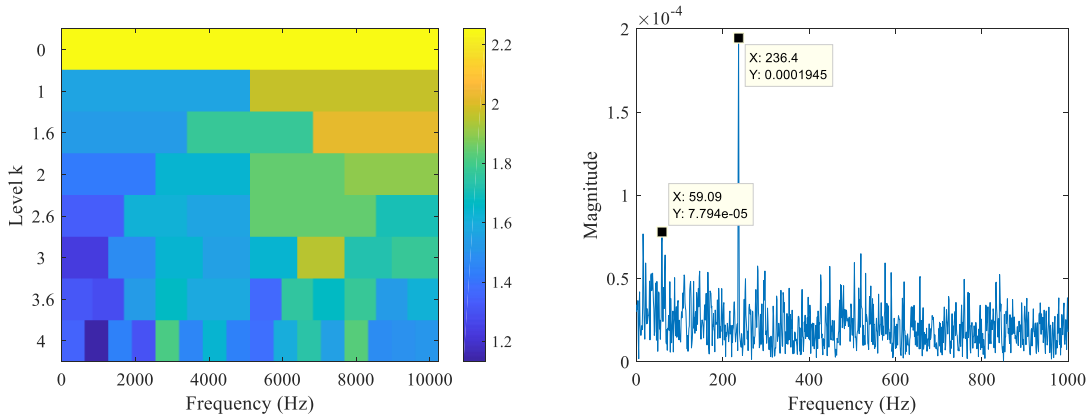


Figure 47 – (a) SE Infogram (b) Frequency Domain of the Filtered Signal by the SE Infogram

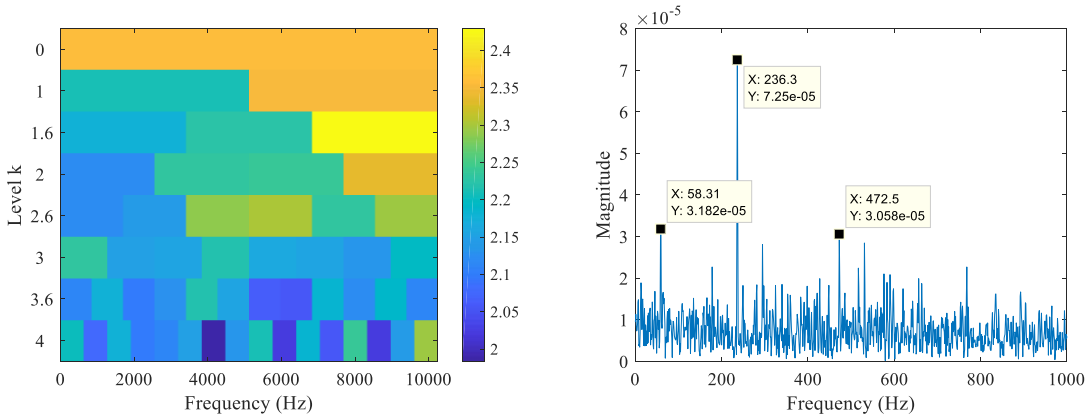


Figure 48 – (a) SES Infogram (b) Frequency Domain of the Filtered Signal by the SES Infogram

The dictionaries are initiated randomly with 50 atoms of which the length is 20. Because larger kurtosis value, in general, enhances the detectability of the fault signature, the artificial kurtosis used as the predicted observation of the UKF is extrapolated in a similar way as in [120]. The value of the artificial kurtosis is show in Figure 49. The kurtosis value combined with the Infogram filtered signal are used as input for the UKF to update the dictionaries.

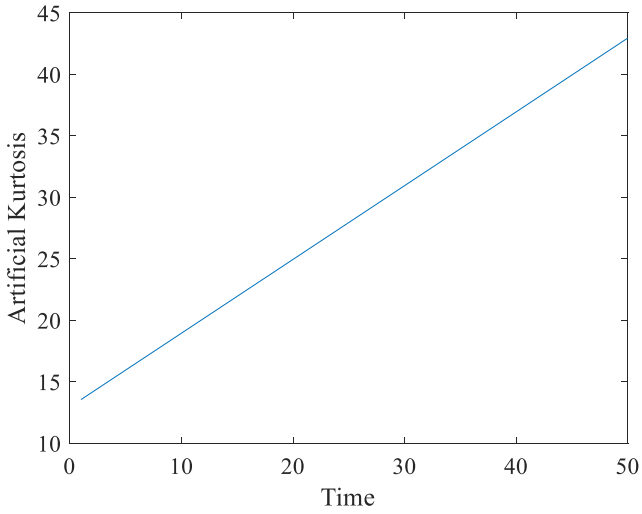


Figure 49 – Artificially Created Kurtosis

The UKF parameter is set as $\alpha = 0.001, \kappa = 0, \beta = 2$ based on the selection criteria proposed in [128]. The randomly initiated dictionary learning diagnosis result is shown in Figure 50 (a). It can be observed that the random dictionary does not provide any information of the defect present in the bearing. However, after the training of the dictionary, it starts to learn the defect frequency and its harmonics as shown in Figure 50 (b). In theory, with more training data available, the performance of the dictionary can be further enhanced. Instead of using the traditional K-SVD method to calculate dictionary. Because the UKF has the advantages of less computation complexity, ability of capturing nonlinearity, and self-adaptation, it could be a viable tool in the calculation of the dictionaries.

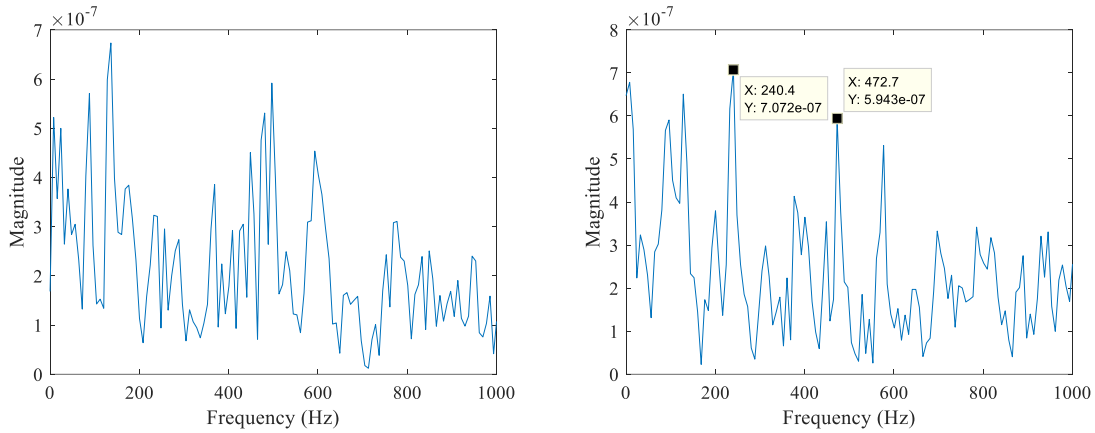


Figure 50 – (a) Initial Dictionary (b) Updated Dictionary

To prove that the proposed dictionary updating algorithm scientific contribution, the initial dictionary weights and the updated dictionary weights are partially shown in Table 4. The total number of dictionaries is selected to be 50 by trial and error. The initial weights θ_k are generated by the normal random number generator. The updated weights θ_{k+1} is in a nonlinear manner. By examining Figure 50, the updated weight yields a better recovery of the fault signature. In addition, it can be observed that some of the weight of the dictionary changes drastically because of the added nonlinearity, and the difference of magnitude could be significant. Rather than the traditional linear method such as the recursive least square algorithm. The weight update of the dictionaries is completed using less iterations and possibly shorter computation time. One interesting thing to study in the future research would be the examination of elimination of the dictionary with the small weight. In this way, a possible structural adaptation could be achieved.

Table 4 – Comparison of Dictionary Weights Before and After Training

Dictionary Number	Initial Weight	Updated Weight
1	0.724	-2.655
2	0.345	1.857
3	0.247	3.013
4	-1.357	7.739
5	1.664	0.173
6	-0.140	2.890
7	-1.266	-6.875

To further prove the proposed method, a widely used empirical mode decomposition (EMD) method is tested for comparison. Different intrinsic mode functions (IMFs) of the decomposed signal are examined, and the frequency plots are shown in Figure 51. The Hilbert spectrum of the vibration signal is plotted in Figure 52. The Because of the EMD method sifting process is similar to a bandpass filtering process, only the first 4 IMFs are shown.

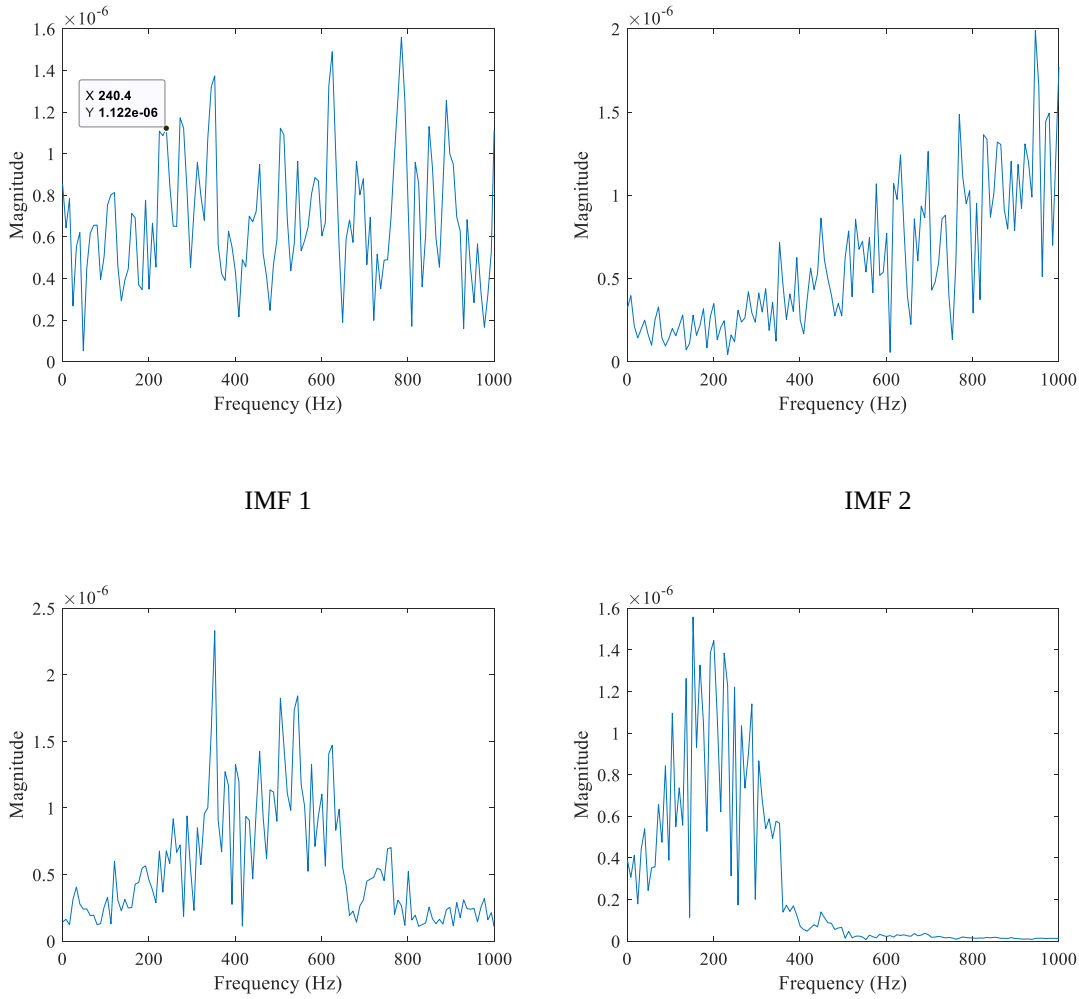


Figure 51 – Frequency Spectrum of IMF 1-4

The traditional EMD method is incapable of detecting the fault signal. By examining the Hilbert spectrum in Figure 52, the EMD method is only sensitive to the high frequency spectrum in the 3000 Hz to 7000 Hz range, which may be beneficial for high frequency resonance analysis. However, for rolling element bearing operating at a relative low speed, the lower frequency components contain more direct information indicating the health state of the machinery.

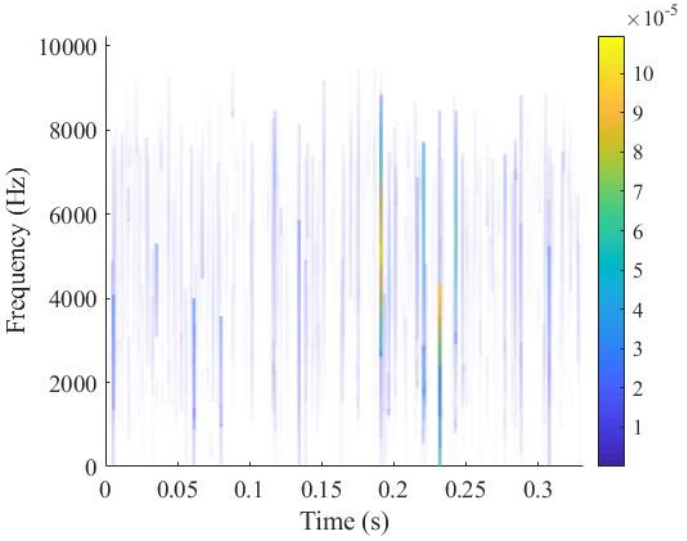


Figure 52 – Frequency Spectrum of IMF 3

2.4.4 Conclusions

In this section, a nonlinear adaptive dictionary learning algorithm for bearing diagnosis is proposed. The proposed model uses the SE and SES Infogram with added capability to analyze cyclostationary signal as a preprocessing tool to acquire the training data for the dictionary. The filtered signal contains the bearing fault frequency and its harmonics buried in the system noise. A nonlinear map is created between the artificially created kurtosis and the weights of the dictionaries. The UKF is implemented to update the weights of a randomly initialized dictionary based on artificially created kurtosis and the filtered data from the Infogram. The updated dictionary is used for bearing fault diagnosis. The UKF is computationally cheaper and easier to implement rather than using the traditional K-SVD method in obtaining the dictionary. The proposed model also gives the self-adaptation capability for the dictionary and can be used in online monitoring process. The dictionary learning diagnostic model creates a frame work for the adaptive monitoring of the bearing degradation process. It is more robust and flexible than a fixed parameter

model such as the wavelet denoising and the Kurtogram. An interesting topic to be investigated, as pointed out at the end, is achieving structural adaptation of the dictionary by eliminate the dictionaries with small weight values. By implementing the structural adaptation, computation effort is guaranteed to be reduced. The performance will need to be evaluated in comparison to the non-structural adaptation model.

2.5 CEEMD-Assisted Bearing Degradation Assessment Using Tight Clustering

2.5.1 Introduction

Rolling element bearing is a critical component that ensures the smooth operation of various machineries. To guarantee the continuous operation without service interruption, the accurate knowledge of the health conditions of the bearings is necessary [103, 131, 132]. A wide variety of research has been dedicated into the condition monitoring of the bearing elements in rotating machinery to improve early fault detection, categorization of fault location, and predict the remaining useful life [61, 104, 133-136]. One of the most effective approaches to diagnose bearing failure and forecast bearing performance is the condition-based maintenance (CBM) [137]. The main idea of CBM is to monitor machine conditions continuously and make proper adjustment based on the measured information from the machines [138]. The most popular measurement from the rolling element bearing is the vibration signal since it is easy to acquire and representative of various fault signatures [1]. Other methods such as acoustic emission, temperature monitoring, oil debris analysis, and stator current monitoring are also applied in machinery condition monitoring [139-143].

The general guideline for bearing diagnosis proposed by Randall and Antoni [5] outlines the procedure to perform diagnose as pre-processing, enhancement of signal, and analysis of fault frequency. The pre-processing aims to reduce the random noise that masks the relative weak bearing signals. The most common method implemented is the linear predictor, autoregressive model, to separate out the deterministic and residual part of the acquired signal [144]. The linear predictor is simple to implement, however, the signal separation is sensitive to the model parameter and types of signal. Adaptive noise cancellation is a different method that uses a reference signal to separate the original signal into two uncorrelated components [6, 145]. The advantage of this technique is its capability to adjust its parameter. But, the reference signal with distinctive fault signature may be hard to obtain under certain circumstances. Some vibration data are acquired with inconstant rotating speed of the bearings, which complicate the deterministic fault frequency detection. One way to alleviate this problem is through resampling techniques [113]. The time synchronous averaging method significantly reduces the effect of variable rotating speed during the data acquisition [146]. The vibration data is generally resampled using constant angular increments which requires the angular increments to be relatively small [111]. Although, most recent research show success on large angle variation without using angular measurement data [13], the complexity and narrow application prevent the method from being widely implemented.

After the signal preprocessing, various signal to noise ratio enhancement methods are implemented to remove other noise. The general techniques are categorized as time-domain, frequency-domain, and time-frequency analysis. The previous two types of techniques are well documented, and the time-frequency analysis is the most widely

implemented nowadays because it combines the benefits of time and frequency domain analysis. Yazici *et al.* implemented an adaptive, statistical short time fourier analysis to diagnose the bearing damage severity based on stator current measurement [147]. However, this type of feature-based selection method does not guarantee a stable categorization. The Kurtogram [51] proposed by Antoni uses wavelet packet transform to separate the signal into various frequency domains. In each of the frequency domains, the kurtosis of the complex envelope of the filtered signal is calculated, and appropriate frequency band of the signal is selected to extract critical information. Although the Kurtogram is effective in most applications, its inability to reveal the cyclostationary characteristics of the signal could result inaccurate diagnosis, which is examined in [14].

Empirical mode decomposition (EMD) [148] is proved to be effective in analyzing nonlinear and non-stationary signal. By implementing the EMD directly, the preprocessing step is not necessary because the EMD method cancels out the white noises and filtered out the unwanted signal. The main drawback of the EMD is the mode mixing. Various techniques have been proposed to alleviate the problem of mode mixing and incapability of processing signal with low signal-to-noise ratio (SNR). The complementary ensemble empirical mode decomposition (CEEMD) is the most up-to-date improvement of the traditional EMD [149]. By adding adaptive noise and performing averaging calculation, the CEEMD further alleviates the mode mixing problem of the previous documented research. The white noise added to the signal is critical in successfully performing the CEEMD. However, no documented literature has proposed a way to calculate the standard deviation of the added noise. In this section, a statistical resampling method is implemented within the CEEMD to calculate the proper standard deviation of the white noise added.

Clustering is an unsupervised learning classification method that has been widely used in multiple engineering applications [150]. During the last decades, a significant amount of variations of the clustering algorithm have been applied to the fault detection of bearings. Yiakopoulos *et al.* implemented the K-means clustering algorithm to classify the fault location of the bearings from measured vibration signal [151]. The method does not require training data from defective bearings as the other ones require. Yuwono *et. al* implemented the particle swarm clustering and Hidden Markov Model (HMM) to detection the fault location of bearings in [152]. The literature provides insights on the effect of model parameters over the diagnostic performance. Other similar clustering algorithms focus on the calculation of features and establishment of different criteria for data classification [153-155]. None of the documented literature has address the issue of repeatability and stability of the cluster algorithm. For instance, several data point may not be classified into the same group during different classification process. The boundary of classification remains ambiguous. In this section, we proposed using a tight clustering algorithm to stabilize the ambiguous data point through a repetitive sampling method.

Tseng and Wong [156] proposed the tight clustering method to identifying the tight and stable patters in the data of unsupervised clustering learning without prior knowledge of the number of clusters. Tight clustering has been widely used in cluster analysis on biological applications, such as microarray analysis and gene expression data analysis [157]. Maitra et al. [158] provided an initialization approaches for deciding the number of clusters as the initials values for the clustering methods. Gaussian mixture model, as a popular clustering method, aims to find a mixture of multidimensional Gaussian distributions that best characterize the input data. Comparing to the K-means method for

clustering data, Gaussian mixture model provides a soft cluster assignment that quantify the uncertainty of the clustering processes. Maitra and Melnykov [159] studied the performance of the Gaussian mixture model clustering proposed by Fraley and Raftery [160]when different initialization strategies were adopted. The literature on using tight clustering on Guassian mixture model for application in bearing degradation assessment is still lack.

In this section, a bearing diagnostic model is proposed using the CEEMD and combined Gaussian mixture model with tight clustering algorithm to assess the degradation stage of the bearing element. The vibration data is initially segmented into small blocks. Then the bootstrap resampling method is implemented, and the optimal input value of the noise standard deviation for the CEEMD algorithm is obtained. After obtaining the input parameters of the CEEMD, the vibration signal is decomposed, and the intrinsic mode function (IMF) containing the fault features are analyzed. Various features carrying important diagnostic information obtained from the analysis of the IMF are calculated. The tight clustering algorithm is implemented to distinguish the bearing degradation stage based on the extracted feature space. With the collection of run-to-failure experimental data of bearings, the degradation stage of the bearings is classified as needed. The proposed algorithm can be implemented in rotary machines through various industry sectors to evaluate the damage severity of the machines and facilitate the maintenance of the machines through their service lives.

2.5.2 Diagnostic Model

The proposed bearing diagnostic model is shown in Figure 53. The acquired vibration signal of the bearing is initially randomly sampled to reduce computation time. The sampled data is reassembly and divided into small blocks. The bootstrap resampling technique is used in each of the blocks with random permutation and multi iterations to obtain the noise standard deviation required for the input of the CEEMD. The CEEMD decomposes the sampled data into various IMFs. The IMFs are examined to select the IMF containing the fault signature based on the known calculated fault frequency. The distinctive vibration feature space is calculated from the selected IMF. After obtaining the feature space, the tight clustering algorithm is implemented to determine the degradation severity from the data obtained at different health stages of the bearing. An experimental bearing vibration data set carrying various degradation stages information is used to validate the purposed model.

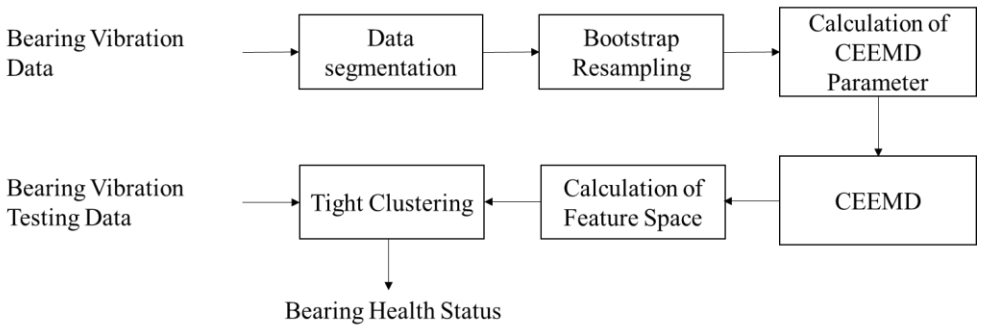


Figure 53 – Unsupervised Bearing Diagnostic Model

2.5.2.1 Optimized CEEMD

The ensemble empirical mode decomposition (EEMD) is an effective tool to analyze machinery vibration signal. However, the problem of mode mixing caused by noise could lead to faulty diagnostic information. To alleviate this problem and improve the

performance of the mode decomposition, Torres *et al.* purposed adding a particular noise to each stage of the decomposition to counteract the existing noise in the acquired signal [161]. Although the method shows improvement over the traditional method, no specific guideline of the added noise parameter selection is proposed. Therefore, to improve the performance of the CEEMD, a bootstrap based resampling algorithm is proposed to determine the optimal parameter of the added white noise.

The CEEMD algorithm implemented here is based on the traditional CEEMD algorithm. It assumes the vibration signal can be represented as:

$$y(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \tag{103}$$

where n is the total number of the IMFs decomposed using the traditional EEMD [162], r_n is the n-th residue. The IMFs satisfies the criteria of equal amounts of extrema, and the envelope containing the extrema is zero. The EMD is an iterative process until the residue term cannot be further decomposed. The first residue can be denoted as:

$$r_1(t) = y(t) - IMF_1(t) \tag{104}$$

where r_1 contains the data of the original vibration data minus the first IMF. The following IMFs can be obtained by following:

$$r_n(t) = r_{n-1}(t) - IMF_n(t) \tag{105}$$

The CEEMD modifies the input to the EMD as:

$$y^*(t) = y(t) + \varepsilon_0 w_n(t) \tag{106}$$

where ε_0 is the noise standard deviation, which is critical for the CEEMD algorithm. w_n is the Gaussian white noise. The first mode by CEEMD, IMF_1^* , is computed as:

$$IMF_1^*(t) = \frac{1}{I}\sum_{i=1}^{n} IMF_1^i(t) = IMF_1(t) \tag{107}$$

where I is the number of realization, IMF_1^i is the intermediately calculated IMF, IMF_1 is the IMF obtained by traditional EEMD. The residue in this case is represented similar to the one shown in (108):

$$r_1(t) = y(t) - IMF_1^*(t) \tag{108}$$

Afterwards, the residue r_1 is modified as:

$$r_1^*(t) = r_1(t) + \varepsilon_1 E_1(w_n(t)) \tag{109}$$

where E_1 denote the EMD operator for the first mode. I times of realization is executed until the second mode is obtained as:

$$IMF_2^*(t) = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(t) + \varepsilon_1 E_1(w_i(t))) \tag{110}$$

The *n*-th residue is represented as:

$$r_n^*(t) = r_{n-1}^*(t) - IMF_n^* \tag{111}$$

Then the following modes are obtained similarly as shown in (110). Once the residue cannot be further decomposed, the original signal can be expressed as:

$$y(t) = \sum_{i=1}^{n} IMF_i^*(t) + r_n^*(t) \tag{112}$$

As can be observed from (106) and (107), the standard deviation of the noise and the number of realizations are two important input parameters for the CEEMD. The number of realizations is not as critical once it passes a threshold value, and it can be found by trial and error. However, no documented research has proposed a way to calculate the standard deviation of the added noise. A bootstrap resampling method is proposed in this section to provide a possible method to obtain the standard deviation. A sample size and number of iterations are determined by trial and error. The actual sample number should be determined based on the sampling frequency. The assumption of the bootstrap method assumes the signal is stationary within each sample. Therefore, if the sample size is too large for the bearing degradation data, the signal within a sample may not satisfy the assumption of stationary. After the sample number and iteration number are determined, the first sample is selected. Then this sample is permuted for multiple times based on the selected number of iterations. The differences between the observations of the first sample and the permuted samples are calculated. The standard deviation of the differences is then

obtained. This process is continued until all the data is sampled. Finally, the mean value of the standard deviations is selected to be the input parameter for the CEEMD.

2.5.2.2 Feature Calculation

The success implementation of the unsupervised learning is hinged on the feature selection. The time domain statistical data generally indicates the machine health conditions [163]. Some common features include the RMS value, kurtosis, skewness, peak-to-peak value, crest factor [61, 163, 164]. The general features and corresponding equations are shown in Table 5 where x_i represent the sample point from the acquired vibration signal, n is the total number of sample point.

Table 5 – Time domain features

Feature	**Equation**
Root mean square	$P_1 = \sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n}}$
Kurtosis	$P_2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^4}{(n-1)\sigma^4}$
Skewness	$P_3 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^3}{(n-1)\sigma^3}$
Peak-to-peak value	$P_4 = \max(x_i) - \min(x_i)$

Feature	**Equation**
Crest factor	$P_5 = \frac{\max\left\|x_i\right\|}{P_1}$
Standard deviation	$P_6 = \sqrt{\frac{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}{n-1}} = \sigma$
Square root average	$P_7 = \frac{\sum_{i=1}^{n}\sqrt{x_i}}{n}$
Impulse factor	$P_8 = \frac{\max\left\|x_i\right\|}{\left(\frac{1}{n}\sum_{i=1}^{n}\left\|x_i\right\|\right)^2}$
Shape factor	$P_9 = \frac{P_1}{\frac{1}{n}\sum_{i=1}^{n}\left\|x_i\right\|}$
Latitude factor	$P_{10} = \frac{\max\left\|x_i\right\|}{\frac{1}{n}\sum_{i=1}^{n}\left\|x_i\right\|}$
Margin factor	$P_{11} = \frac{\max\left\|x_i\right\|}{\left(\frac{1}{n}\sum_{i=1}^{n}\sqrt{\left\|x_i\right\|}\right)^2}$
Kurtosis factor	$P_{12} = \frac{P_2}{\left(P_1\right)^4}$

Other than the time domain feature, two frequency domain features are calculated. The signal magnitude around the fault frequency is obtained by performing the Hilbert transform and summing up the magnitude of the frequencies centered at the fault frequency. In this case, a interval of 10 Hz around the fault frequency is selected to account for the possible slippage in the bearing components and difference between the theoretical calculation and actual fault frequency. Another frequency domain feature, Negentropy, is also computed and used as a feature for the tight clustering algorithm. The envelope spectrum of the data is first obtained [48]. Then the negentropy is calculated as [48]:

$$N = \left\langle \frac{e_i^{\,2}(n; f, \Delta f)}{\left\langle e_i^{\,2}(n; f, \Delta f) \right\rangle} \ln \frac{e_i^{\,2}(n; f, \Delta f)}{\left\langle e_i^{\,2}(n; f, \Delta f) \right\rangle} \right\rangle \tag{113}$$

where $e_i(n; f, \Delta f)$ is the instantaneous flow of energy in the band centered on f with bandwidth Δf. After the time domain and frequency domain features are obtained, tight clustering algorithm uses the features as input to determine the degradation stage of the bearing.

2.5.2.3 Tight Gaussian Mixture Model Clustering

Tight clustering is a resampling based method in the clustering analysis. Given the features extracted in the previous section, the representative data points or time periods that well represent the different bearing health condition are determined. Most of the clustering algorithms aim to give every data point a label so that all data points are assigned to one or more clusters. For bearing health condition assessment, there are points belonging to

transition phase and we are mainly interested in finding the most representative and tight points for bearing health condition identification.

Therefore, tight Gaussian mixture model clustering method is developed for stable bearing health condition clustering. The tight Gaussian mixture model clustering method leverages the advantage of finding the tight clusters in clustering [156] and Gaussian mixture model for model-based clustering, which does require the prior knowledge of the number of clusters.

The algorithm of tight Gaussian mixture model can be summarized below.

(1) Randomly sampling from the original data matrix Y and denoted as Y_s;
(2) Apply the Gaussian Mixture Model (GMM) for clustering on data X_s to obtain the k cluster centers $C(Y_s,k)$, where k is decided by GMM.
(3) Implement the supervised learning classification on the original data Y based on the label information (cluster center $C(Y_s,k)$) to get a comembership matrix $D[C(Y_s,k),X]$ (see Tibshirani 2005) [165]
(4) Repeat Step (1)-(3) independently B times. Take the average of the comembership matrix $\bar{D} = \frac{1}{B}\sum_{b=1}^{B} D^{(b)}$.
(5) Find the candidates of tight clusters by the hard threshold $\bar{D}_{ij} \geq \delta$ a fixed constant for $\forall i, j$.

2.5.3 Experimental Result

To examine the proposed diagnostic model, the experimental data from Qiu *et al.* [16] is used to validate the clustering algorithm. The overall acquired bearing vibration signal is shown in Figure 54. The measured acceleration of the bearing at the initial and intermediate stage is relative constant. However, the acceleration increases exponentially around 15000 min, which indicates the bearing is heavily damaged. The boundary of healthy and light damaged bearing is difficult to quantify from the acquired signal.

Therefore, the pre-processing of data and the calculation of features in the time and frequency domain become critical to extract the health state of the bearing.

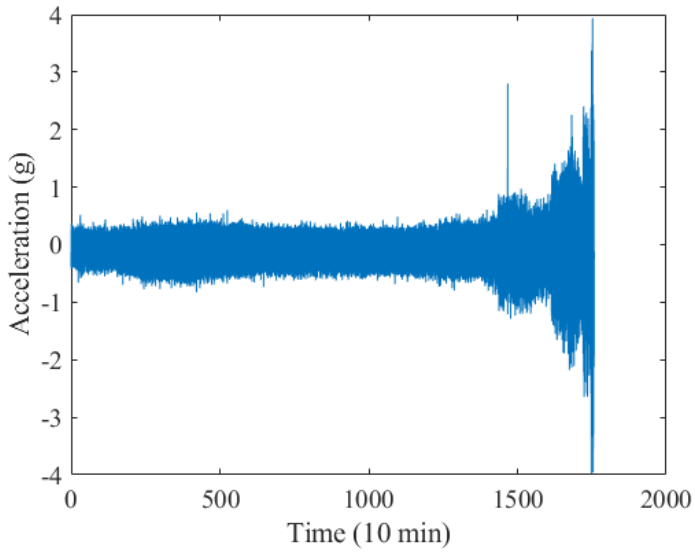


Figure 54 – Acquired Bearing Vibration Signal

To reduce the effect of noise, the improved CEEMD is implemented to extract the IMF which contains the most information related to bearing health. The CEEMD parameter is selected using the bootstrap resampling method, and the noise standard deviations for the vertical and horizontal data are 0.0797 and 0.0856 respectively. The number of realizations is set to be 300 by trial and error because that once the number of realizations passes a threshold value, the effect of the number is negligible. After the CEEMD process is completed, the IMFs are examined. In this case, the IMF 6, which contains the fault frequency signal, is selected for further analysis. The vertical and horizontal signal of the IMF 6 are plotted in Figure 55

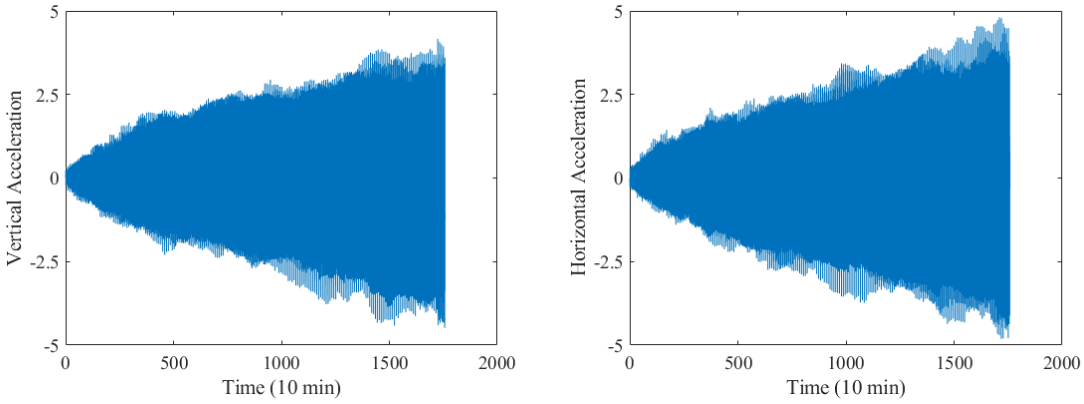


Figure 55 – Vertical and Horizontal Acceleration after CEEMD processing

The degradation trend is clearly revealed after the CEEMD processing of the noisy vibration signal. The data is further split into small blocks of length 2500 which is selected to ensure the frequency domain features are revealed. The 13 different time and frequency domain features from the vertical and horizontal data are calculated, and 70% of the data is used for training and 30% of the data is used for testing resulting assembled feature matrices of dimension 280 by 26 and 120 by 26. A random number generator is used to make sure an even spread of data is selected. The categorization calculated by the tight clustering algorithm is shown in Figure 56. Because of the nature of the algorithm, only the most representative data are shown in the figure. The data between 70 to 220 are the transitional data which oscillate between different categories. Thus those data are eliminated by the algorithm. Because the data point before 250 already indicated heavy damage of the bearing, it is unnecessary to use the data after 250 to assess the health condition. Therefore, the data point after sample 250 are eliminated.

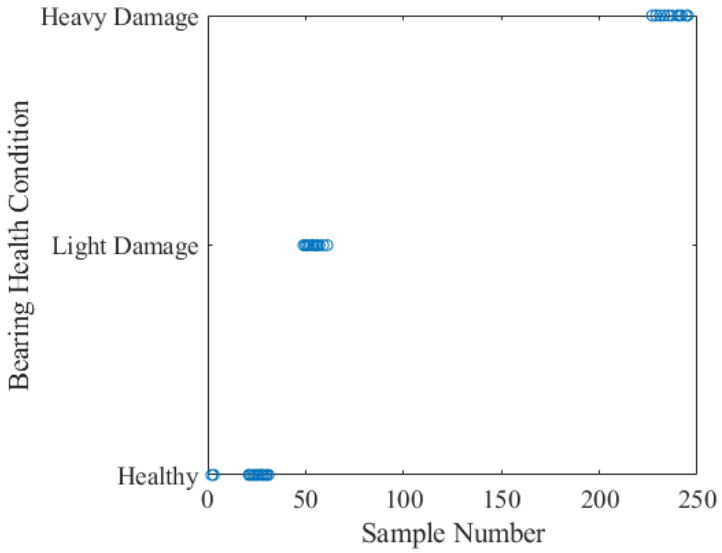


Figure 56 – Tight Clustering Training Data

The result of tight clustering using testing data is calculated by the K-nearest neighbor algorithm, which determines the Euclidean distance between each data point. The testing feature point that has the minimum distance between the training feature point is categorized as the same category. The 120 test data points with corresponding classification are plotted in Figure 57. It can be observed that only one of the testing data is misclassified using the proposed tight clustering algorithm. The test accuracy is calculated to be 99.17%

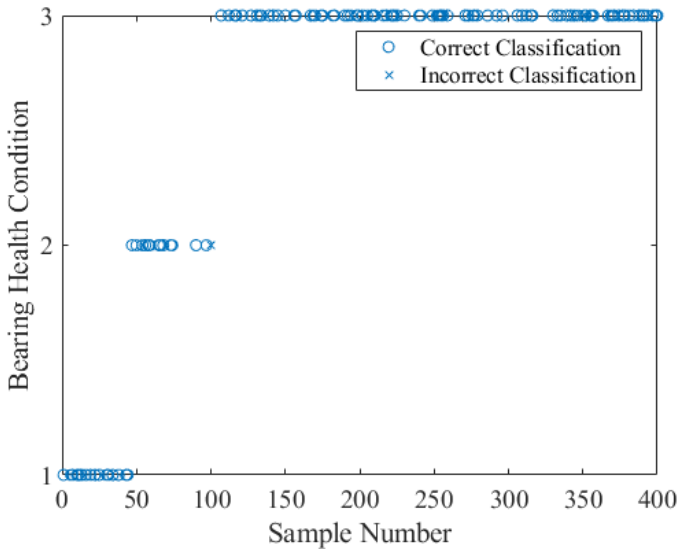


Figure 57 – Tight Clustering Testing Data

To further illustrate that the tight clustering algorithm facilitates the training process by using less quantity and more accurate data, a standard GMM based clustering algorithm is used as a benchmark for comparison. The training data used by the GMM algorithm is shown in Figure 58. Without the tight clustering algorithm, the GMM model uses all the

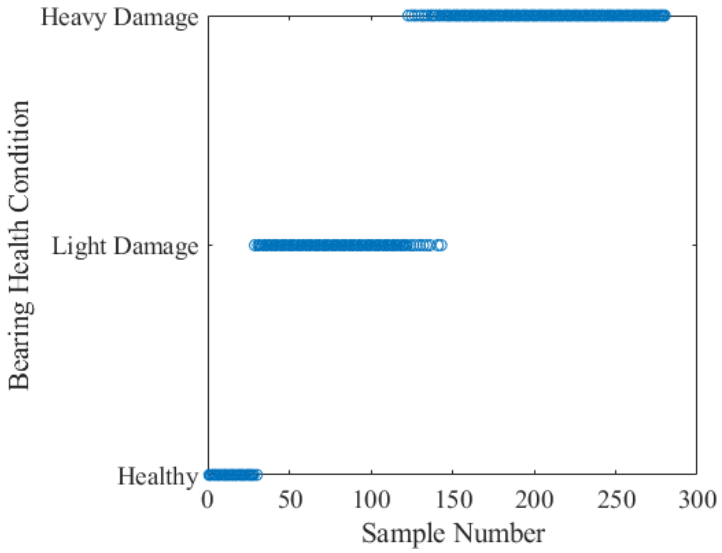


Figure 58 – GMM Clustering Training Data

training data to categorize the bearing healthy condition, which consumes more time for the training process. The test result is shown in Figure 59. Comparing Figure 57 and Figure 59, one can observed that same amount of data is categorized correctly under the healthy condition for both models. However, the GMM categorizes more of the data into the light damage category and less of the data into the heavy damage category, which could result underestimation of the current damage level existed in the machine. In addition, more data overlap is observed between the heavy damage data and light damage data, which are resulted from the unstably categorized training data of the traditional GMM.

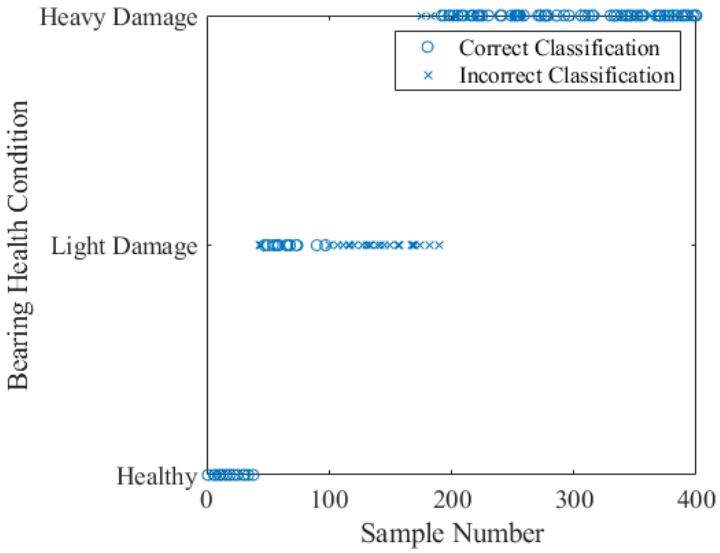


Figure 59 – GMM Clustering Testing Data

The comparison of the two method is shown in Table 6. By using the proposed tight clustering algorithm, the training data length is shortened 87.5% which significantly improve the training effectiveness. In addition, a significant improvement of the classification accuracy is achieved by using more representative data generated by the proposed algorithm.

Table 6 – Comparison Result

Method	Training Data Length	Accuracy
Tight Clustering	35	99.17%
GMM Clustering	280	78.33%

2.5.4 Conclusions

In this section, a degradation assessment model by implementing the improved CEEMD and combined Gaussian mixture model with tight clustering algorithm is proposed. The model uses the improved CEEMD algorithm as a pre-processing tool to filter the noisy signal and extract fault features in the time and frequency domains from the bearing vibration data. The tight clustering algorithm is implemented to improve the training efficiency and accuracy. In this diagnostic model, an approach to calculate the input parameter for the CEEMD using the bootstrap resampling method is proposed to improve the CEEMD's capability of denoising the vibration signal. In addition, by combining the Gaussian mixture model with tight clustering algorithm, more representative data are selected for the training of classifier. The successfully elimination of data with oscillatory categorization provide a more accurate training data set and reduces computation time. The proposed method is implemented in diagnosing the degradation stage of element bearing. The result shows early detection of the bearing health condition based on the calculated vibration fault signature. Less data categorization overlap is observed by comparing the proposed model with the traditional GMM. A significant improvement of 20% accuracy is observed for diagnosis of the health condition of rolling element bearings. Although the proposed method has only been applied to rolling element bearing with outer race failure, theoretically the algorithm can be extended for wider application areas. For instance, by changing the CEEMD decomposition level, features related to other fault signatures are extracted to evaluate the degradation of bearings. In addition, gene clustering has been implemented in medical industries. The applicability to other types of failure will be evaluated in the future studies.

2.6 Bayesian Optimized Deep Convolutional Network for Bearing Diagnosis

2.6.1 Introduction

Under normal degradation process, the intensity of the vibration of bearings remains constant for an extended duration or even decreases because of the edge rounding effect for the sharp defect [166]. When a significant amount of damage has accumulated, the defect growth becomes unstable [167]. Hence, the intensity of vibration increases drastically. At this point, the bearing is approaching the end of its life, and maintenance is immediately required. Ideally, the maintenance should be performed before crack within the bearing elements grows unstable. However, because of the nature of the vibration signal acquired from bearings, the defect growth is difficult to estimate in the early and medium degradation stages without any advanced signal processing techniques. Additionally, harsh operating conditions are often encountered during actual applications, which drives up the difficulties to have an accurate estimation of bearing defect [168]. The abovementioned factors contribute to the complexity of bearing diagnosis, which regularly scheduled maintenance cannot satisfy the industrial need. Therefore, the concept of condition-based-maintenance (CBM) is proposed to compensate for the lack of flexibility and accuracy of the traditionally accepted maintenance schedule [6, 73, 169]. The CBM proposes using sensors to monitor the bearing degradation online instead of measuring the vibration signal in a scheduled interval [137]. The online monitoring is capable of tracking the bearing behavior based on its usage and historical data. Because the online monitoring relies on vibration signal to predict the health status of bearings, the capability of early detection of bearing fault and an accurate correlation between the vibration signal and fault size are critical for successful implementations of the CBM.

To implement the diagnosis part of the CBM, the general approach includes data acquisition, data processing, identifications of fault locations, evaluation of fault size, and decision making [5]. The accelerometer measurements, acoustic emissions are the most accepted nowadays for the diagnosis and prognosis of rotating machineries because of the good correlation with the defect size and ability to detect underlying crack [7, 170]. In addition, magnetic plugs and thermocouples are used as auxiliary devices to assist the measurement of defect severity. In comparison with acoustic emission, accelerometers are easier to install, cheaper and require much lower sampling frequency. Therefore, measurement from accelerometers is more popular to be used for the diagnosis of bearings.

After the acquisition of the vibration signal, various signal processing techniques are implemented to enhance the signal to noise ratio and extract critical features for diagnosis. The obtained information is used for classification for fault type or determination of fault size. A decision is made after analysis of the obtained and derived information. The data acquisition involves sensors and data acquisition system. The selection of devices is based on budget and types of signal to be measured. Piezoelectric or MEMS accelerometers are the most common types of vibration monitoring sensors [48, 171].

Piezoelectric sensor has better signal to noise ratio, but it is more expensive and requires signal amplification. MEMS sensor is cheap and signal processing unit can be built into the circuits, however, the signal to noise ratio is generally lower than the Piezoelectric sensor. The pre-processing aims to reduce the random noise that masks the relative weak bearing signals. The most common method implemented is the autoregressive model [116], a type of linear predictor. The model separates out the

deterministic and residual part of the acquired signal [75]. Thus, the enhancement of signal to noise ratio is achieved. However, model parameters and types of signal could affect the signal and noise separation for autoregressive models.

Adaptive noise cancellation uses a reference signal, normally a known signal carries the desirable vibration frequencies, to separate the acquired signal into two uncorrelated components [6, 145]. The advantage of this technique is its self-adaptive parameters. But, the reference signal with a distinctive fault signature could be difficult to obtain since the complex interaction of bearing components is generally unknown. In a certain scenario, the acquired vibration signals consist of data under variable rotating speed [168]. To alleviate problem of variable rotating speed, resampling techniques are proposed [172, 173]. With data resampled using constant angular increments, the time synchronous averaging (TSA) method significantly reduces the effect of variable rotating speed during the data acquisition [146]. However, to implement TSA, either the angular velocity needs to be monitored at the same time, or the angular increments are relatively small [174].

After the preprocessing of the signal, various algorithms are used to extract signal within the frequency bands of interest. The time-frequency analysis is the most effective and widely implemented because it monitors the changes of signal within a frequency band as the degradation process occurs in the time domain. As one of the time-frequency analysis tool, empirical mode decomposition (EMD) [148] has demonstrated to be effective in analyzing bearing vibration signals. Implementation of the EMD eliminates the extra step of preprocessing for vibration signal acquired under constant rotating speed because the EMD utilize white noises to cancel out unwanted signals [162]. The major drawback of the traditional EMD is the mode mixing [175]. To alleviate this problem, Yeh *et al.* proposed

an improved algorithm, complementary ensemble empirical mode decomposition (CEEMD), with noise enhanced data analysis method to eliminate residue noise in the intrinsic mode functions [149]. Torres *et al.* uses a particular noise added to each stage of the CEEMD to analyze an electrocardiogram signal [161]. Although the application of CEEMD shows successful result [176-178], no documented research has shown the parameter tuning of the algorithm. Lu implements a resampling method to calculate the standard deviation to be used in the CEEMD [179]. The proposed CEEMD with resampling method is utilized in this section for data processing.

After data processing, the evaluations of fault size are generally achieved by correlating measured defect size with features calculated from vibration signal [7, 26, 140, 180, 181]. Most of the data to estimate defect size are obtained from acoustic emission sensors. The correlation between defect size and vibration signal still remains as a challenge for researchers to explore. Li proposes using a function of shaft rotating speed, radial load and root mean square value of the vibration signal to estimate the defect size in [7]. The method shows good correlation with the measured defect size during a bearing life test. Therefore, we select the method proposed by Li to calculate the theoretical defect size.

The final step is to classify the damage severity of the bearing. The most popularly implemented methods in recent years are the deep learning method [182]. The benefits of the deep learning model include high accuracy, low requirement of signal to noise of acquired signal, and being able to extract high-level features from complex signals automatically. Gan et al. developed a hierarchical network capable of predicting types of fault and defect severity [183]. The method overcame the ambiguous classification difficulty resulted from traditional neural networks and support vector machines. Jia et al.

proposed a deep learning algorithm that is capable of adaptation and solving a highly non-linear problem in condition monitoring of rotating machinery [184]. The algorithm considers the fault location and defect size from bearing vibration data. In addition, improved accuracy is achieved by the added layers. Guo et al. proposed a deep convolution neural network for bearing fault classification and fault size identification in [185]. The method shows the similar improved result in comparison to [183]. But it employed an adaptive learning rate and implemented more layers of the network. Other similar deep learning implements in bearing diagnosis can be found in [186-188]. Most of the documented research have shown high accuracy of classification. However, it does not indicate if the accuracy belongs to the validation data or testing data, which makes it difficult to decide whether the corresponding model is overfitting or not. In addition, most of the algorithms use the acquired signal directly, which contains unwanted information and prolongs processing time. The effect of using preprocessed data by signal processing method remains unknown.

In this section, we proposed a bearing diagnostic model utilizing the CEEMD for signal preprocessing, the defect size calculation equation for defect estimation and the deep convolutional network for classifying the defect severity. Since most of the documented work input the unprocessed bearing signal into the network directly without removing noises and unwanted frequency band, the learning data could possibly be contaminated. The innovation of our proposed work is the utilization of processed data by the optimized CEEMD algorithm with Bayesian optimized deep convolutional network. The CEEMD is implemented to extract the frequency band containing the fault signature. A clear degradation pattern is obtained after the decomposition. The defect size calculation

equation is used to estimate the defect size based on root mean square value of the signal, normalized radial load and rotational speed. After obtaining the defect size, the Bayesian optimized deep convolution neural network is implemented to categorize the defect severity. The result of the section is organized as follows. Section II discusses the diagnostic model. Section III illustrates the experiment conducted. Section IV shows the result of the diagnostic model. Section V concludes this section and propose possible future improvement.

2.6.2 Diagnostic Model

The training scheme of the diagnostic model implemented in this section is shown in Figure 60. The acquired vibration signal from accelerometer is segmented into small groups. The selection of group size depends on the sampling rate and the bearing degradation rate. The grouped data should meet the assumption that the data within each group should be relatively the same magnitude to perform the bootstrap resampling

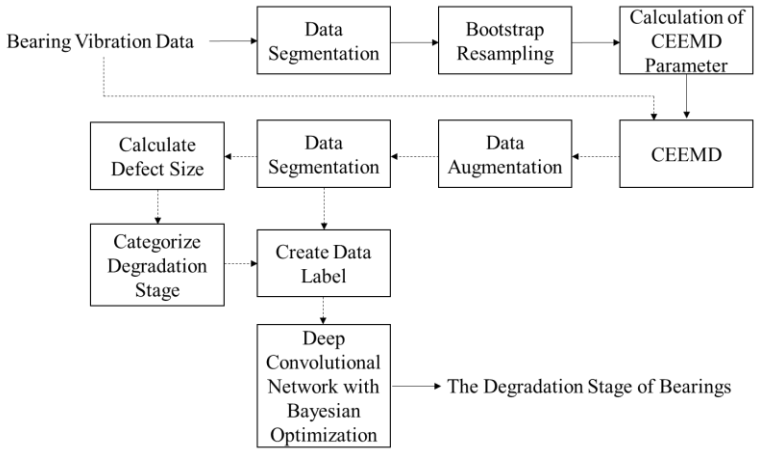


Figure 60 – Proposed Diagnostic Model

algorithm. Once the data within each group have been permutated and resampled. The standard deviation within each group is calculated and recorded. The permutation and resampling are repeated for 1000 times, and the average of the standard deviations are obtained to ensure the accuracy of calculation and prevent outliers. The CEEMD algorithm uses the calculated standard deviation to generate the desirable white noise and decompose the signal into different frequency domains. Each of the decomposed signals is examined by performing Hilbert Transform. The IMF 6 is selected based on the known fault type of the acquired bearing data and calculated fault frequency. After obtaining the preprocessed vibration data, the data is augmented with 40% overlap to facilitate the training of the deep convolutional network. After the augmentation, the data is divided into the small group again, and the experimentally determined fault size calculation equation based on [7] is used to calculate the fault size based on the preprocessed signal. From the calculated fault size, the degradation stage of the bearing can be classified based on the users' requirement. The label is attached to the grouped data for classification. The labeled data is input into the deep convolutional network for training, validation and testing. The following

subsection details out the algorithm used in CEEMD, defect size calculation, and Bayesian optimized deep convolutional network.

2.6.2.1 CEEMD with Bootstrap Resampling

The EMD has been widely implemented to analyze vibration signal from rotating machineries. The mode mixing of the EMD limits its capability of accurate fault detection. To resolve this issue, the CEEMD is developed by using added white noise to cancel the existing noise in the acquired signal. To achieve better noise cancellation, the parameter of the added white noise is critical. A bootstrap based resampling algorithm proposed in [179] is implemented. The CEEMD assumes the vibration signal can be decomposed into multiple intrinsic mode function (IMF) as:

$$y(t) = \sum_{i=1}^{n} IMF_i(t) + r_n(t) \tag{114}$$

where n is the total number of the IMFs decomposed using the EEMD [162], r_n is the residue after n times of decomposition. The IMFs meet the criteria of equal numbers of maxima and minima. The algorithm continues until the residue term shown in equation (115) cannot be further decomposed

$$r_n(t) = r_{n-1}(t) - IMF_n(t) \tag{115}$$

The CEEMD modifies the input to the EMD by adding Gaussian white noise to the original signal as:

$$y^*(t) = y(t) + \varepsilon_0 w_n(t) \tag{116}$$

where ε_0 is the standard deviation of the noise, which is determined by the bootstrap resampling method; w_n is the standard Gaussian white noise. The first mode, IMF_1^*, is computed as:

$$IMF_1^*(t) = \frac{1}{I}\sum_{i=1}^{n} IMF_1^i(t) = IMF_1(t) \tag{117}$$

where I is the number of realizations, IMF_1^i is the intermediately calculated IMF, IMF_1 is the IMF obtained by the EEMD. The residue in this case is represented similar to the one shown in (115):

$$r_1(t) = y(t) - IMF_1^*(t) \tag{118}$$

With the added Gaussian noise, the residue r_1 is modified as:

$$r_1^*(t) = r_1(t) + \varepsilon_1 E_1(w_n(t)) \tag{119}$$

where E_i denotes the EMD operator for the i-th mode, ε_1 is the standard deviation of the added noise. I times of realizations are executed until the second mode is extracted as:

$$IMF_2^*(t) = \frac{1}{I}\sum_{i=1}^{I} E_1(r_1(t) + \varepsilon_1 E_1(w_i(t))) \tag{120}$$

The n-th residue is calculated as:

$$r_n^*(t) = r_{n-1}^*(t) - IMF_n^* \tag{121}$$

Then the following IMFs are obtained similarly as shown in (120). Once the residue cannot be further decomposed, the original signal can be expressed as:

$$y(t) = \sum_{i=1}^{n} IMF_i^*(t) + r_n^*(t) \tag{122}$$

The standard deviation of the added Gaussian noise and the number of realizations is two important parameters for the CEEMD. The decomposition tends to converge once the number of realizations passes a threshold value. The value can be determined by trial and error or select a relatively large number if computation power is allowed. The bootstrap resampling method provides a possible way to obtain the standard deviation of the added noise. The sample size is selected based on the assumption that the as the resampled data block is stationary. The number of iterations is generally a large number than 1000. After the sample size and iteration number are determined, the first sample is permuted for the number of iterations selected. The differences between the observations of the first sample and the permuted samples are calculated. The standard deviation of the differences is obtained for each iteration. The algorithm keeps running until all the samples have been permuted. Finally, the average value of the obtained standard deviations from each sample is selected to be the standard deviation of the added Gaussian noise for the CEEMD.

2.6.2.2 Experimentally Determined Defect Size Calculation

Li demonstrates that the vibration and acoustic emission signals are affected by defect severity, rotating speed, load, temperature, lubrication in [7]. An experimentally

determined defect size calculation equation is proposed based on the shaft rotating speed, radial load and root mean square of the acquired signal. The equation is determined using regression model shown as:

$$y = 0.0912 + 0.495x_1 + 0.188x_1x_2 - 0.0414x_1x_3 \tag{123}$$

where y is the normalized signal root mean square, x_1 is the normalized bearing defect width, x_2 is the normalized shaft rotating speed, and x_3 is the normalized radial load. The negative coefficient indicates that adding radial load will reduce the vibration intensity of the bearings. To use this model, a correction factor is added so that the initial defect size will be zero for the experimental data used in this section. The modified equation adding a correction factor c is shown as:

$$defect\ size = c + \frac{RMS - 0.0912}{0.495 + 0.188 \times speed - 0.0414 \times load} \tag{124}$$

This equation is unit-sensitive, and the unit of the *RMS* is acceleration (*g*), the unit of the *speed* is RPM, and the unit of load is in *lbs*.

2.6.2.3 Deep Convolutional Network with Bayesian Optimization

The deep convolutional network as a novel and powerful tool to capture the complex dependency between input and output signals has been implemented in various diagnostic applications. The convolutional neural network (CNN) works well in identifying patterns in data and has been adopted in various applications in the scenario: absence of human experts and adaption of solutions to cases [189-191] because of its

capability of solving highly nonlinear problems and generalization to different data types. Our 1D CNN architecture consists of two types of layers: convolution and fully connected layer. In the convolution layers, the data is convolved using local learnable kernels to form the output feature maps. The activation function within the network is represented as:

$$x_{out}^{(l)} = g(h^{(l-1)}) = \sigma(\sum x_{in} * W_j + b_j) \tag{125}$$

where $x_{out}^{(l)}$ is the output feature map of the current layer l, x_{in} is the input data, W_j and b_j are the kernel and bias for the current layer. After the convolution layer, the data is pooled using the maxpooling function and input into the next convolution layer.

Once all the convolution processes are finished, the output data are then flattened and fed into the fully connected layer for processing. The added convolution is sensitive to nonlinear function such as sinusoidal and periodic function which benefits the diagnostic application. The softmax function is utilized in the final classification layer to categorize the degradation stage of the bearing. The activation function between each layer is selected to be Rectified Linear Unit (ReLU) [191] for computation efficiency:

$$[\sigma(z)]_j = \max\{z_j, 0\} \tag{126}$$

The segmented CEEMD processed vibration data are selected as the input of the network and the degradation stage based on defect size are selected to be the output.

The proposed network structure in this diagnostic application is shown in Figure 61 and Table 7 and implemented with TensorFlow. The loss function is selected to be categorical cross-entropy and the optimizer is selected to be Adam. The model is trained

on (Intel Xeon E5-2687W v4 CPU with 64 GB RAM and NVIDIA Quadro M6000 GPU) for 100 epoches with batch size of 2000.

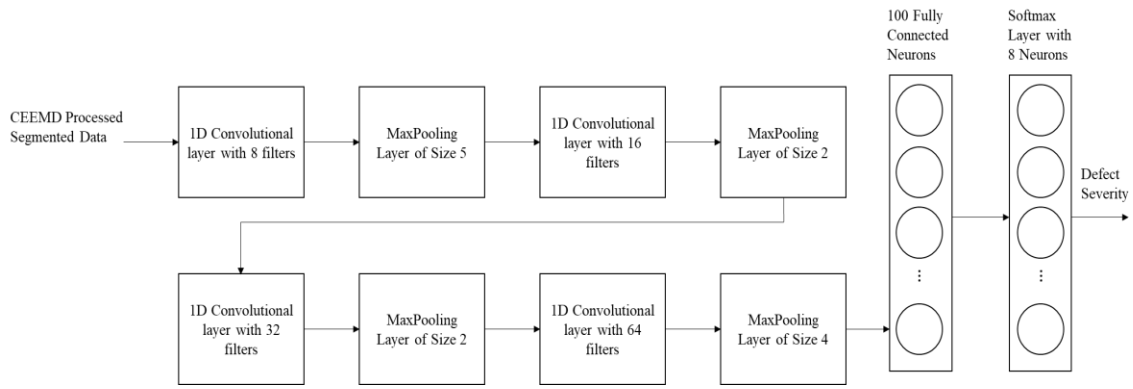


Figure 61 – Deep Convolutional Neural Network

Table 7 – Deep Convolution Network Parameters

# of Layer	Layer Type	Kernel Size	Stride	Filters
1	Convolution 1D	30	2	8
2	MaxPooling 1D	5	5	8
3	Convolution 1D	3	3	16

4	MaxPooling 1D	2	2	16
5	Convolution 1D	3	3	32
6	MaxPooling 1D	2	2	32
7	Convolution 1D	3	3	64
8	MaxPooling 1D	4	4	64
9	Fully Connected	100	N/A	1
10	Softmax Classification	8	N/A	1

The data are split into training, validation and test sets as 60% 20% and 20%. The Kernel size, stride and number of filters are optimized using the Bayesian optimization algorithm described in the following paragraphs.

The Bayesian optimization algorithm facilitate the automatic tuning of the network with boundaries of parameters specified. At each iteration during the optimization, the algorithm utilizes the Bayes' rule to find the minimum or maximum of an objective function *f(x)* within bounded set χ. The Bayesian optimization automatically quantifies the uncertainty of the minimizer or maximizer. The general procedure of the Bayesian optimization can be described as follows. For a generic family of models with data observation *D* and parameter $x \in \chi$, a prior distribution p(x) and a likelihood $p(D \mid x)$ are assumed with given data *D*. We can infer the posterior distribution using the Bayes' rule. The maximizer (or minimizer) of the objective function then follows the maximum a posteriori (MAP) probability:

$$x^* = \arg\min_{x \in \chi} f(x) = \arg\min_{x \in \chi} (x \mid D) \tag{127}$$

In this case, domain χ will be the range of the parameter within the network. We assume the Gaussian process prior such that the observations $D_{1:t} = \{x_{1:t}, y_{1:t}\}$ follows $y_{1:t} \sim Normal(f(x_{1:t}), \sum(x_{1:t}, x_{1:t}))$. Then the posterior probability distribution has the form $f(x) \mid f(x_{1:t}) \sim Normal(\mu_t(x \mid D), \sigma_t^2(x \mid D))$, where $\mu_t(x \mid D)$ and $\sigma_t^2(x \mid D)$ have complicated forms that are usually intractable. The Bayesian optimization establishes a probability model for *f(x)* by selecting various parameter values within the set χ. The model stores the previously calculated *f(x)* value and evaluation the area with higher probability

to generate the minimum or maximum of *f(x)* without relying on the local gradient [192]. Rather than using an exhaustive search algorithm, the Bayesian optimization targets the area with a higher density of lower values of the cost function *f(x)*, which reduce the computation effort significantly. The basic algorithm can be found in [193]. The Bayesian optimization algorithm is combined with the CNN deep learning network to optimize the parameters within the network while reducing the training MSE.

2.6.3 Experimental Setup

The experimental data are obtained from Qiu *et al*.[16]. The vibration data measured in the vertical position of the bearing 1 with failure of its outer race is used to test the proposed diagnostic model. The unprocessed signal is shown in Figure 62. It can be observed from the figure that the initial vibration intensity stays constant for a long period of time. Then, when the crack grows unstable, the vibration intensity increases drastically. The accuracy of bearing diagnosis relies on the early detection of fault. Without any preprocessing method, the reliability of the diagnostic result remains low.

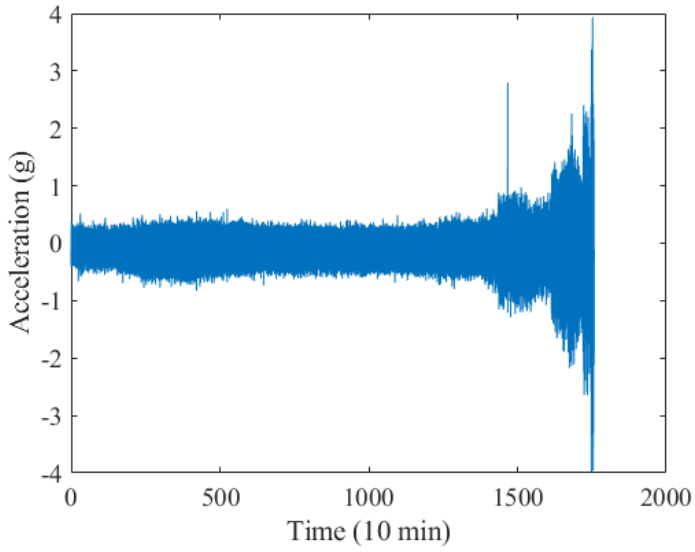


Figure 62 – Bearing Lifetime Test Data

2.6.4 Result

The CEEMD decomposed signal is shown in Figure 63. Compared with Figure 62, Figure 63 clearly shows a degradation growth with respect to time. By monitoring the vibration change of the preprocessed signal, the defect growth of bearing elements can be calculated in the early stage. The data are then augmented with 40% overlap and the defect

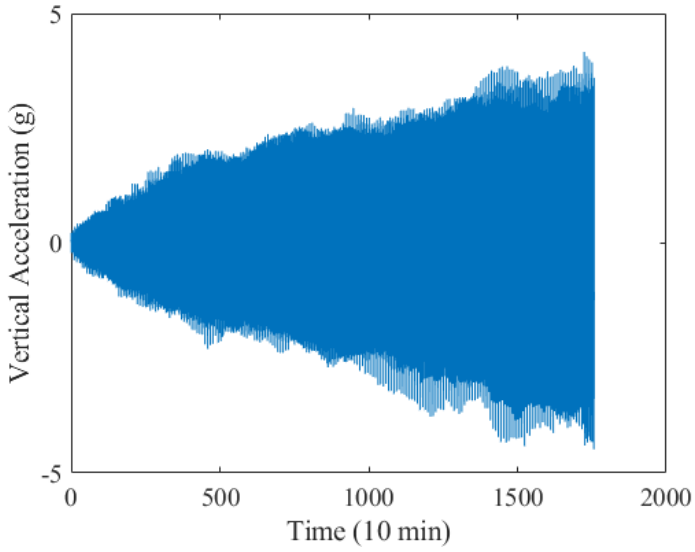


Figure 63 – CEEMD IMF 6

size calculated using Equation (124) is shown in Figure 64. Because the data has been augmented, the unit of the time axis is ignored. It can be observed from Figure 64 that the defect growth is steady in the early stage. At the later stage, the defect growth starts to increase significantly as the slope of the curve increases. The degradation stage is categorized based on the user requirement. Finer classification results in a better capture of the defect size. However, it increases the model complexity and computation cost.

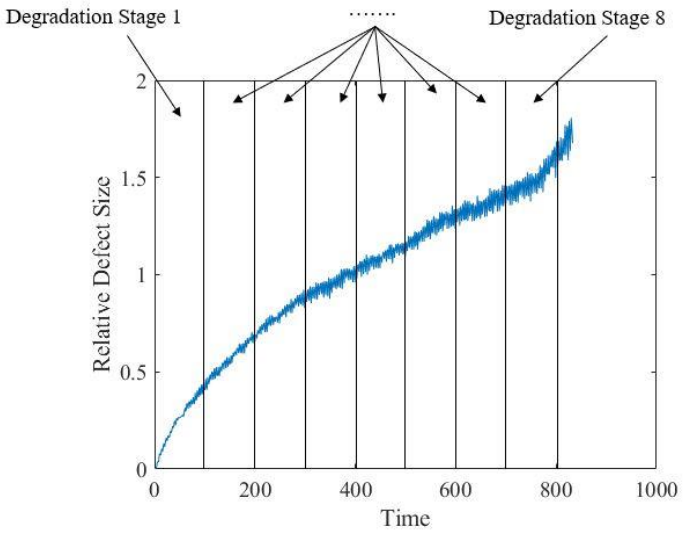


Figure 64 – Degradation Stages of the Experimented Bearing

The selection of the parameter range of the proposed convolutional network is implemented by referring to previously documented network structures. The range is specified in Table 8. Some of the combinations of the parameters may not be achievable in this specific case because of the requirement of layer dimensionality of the convolutional neural network. Those cases are excluded from the test.

Table 8 – Bayesian Optimization Parameter Range

No. Layers	Layer Type	Kernel Size	Stride	Filters
1	Convolution 1D	1-40 with 5 increment	1-3 with 1 increment	1-8 with 2 increment

2	MaxPooling 1D	1-5 with increment	1	1-5 with 1 increment	with	1	N/A	
3	Convolution 1D	1-5 with increment	1	1-3 with increment	1	8-16 with increment		4
4	MaxPooling 1D	1-3 with increment	1	1-3 with increment	1	N/A		
5	Convolution 1D	1-3 with increment	1	1-3 with increment	1	16-32 with increment		4
6	MaxPooling 1D	1-3 with increment	1	1-3 with increment	1	N/A		
7	Convolution 1D	1-3 with increment	1	1-3 with increment	1	32-64 with increment		8

8	MaxPooling 1D	1-4 increment	with	1	1-4 increment	with	1	N/A

Through the Bayesian optimization algorithm, the best parameter selected is shown in Table 7. The validation error and testing error percentage for 5 trials are shown in Table 9. Benchmark models using the network structure proposed by Zhang *et al.* [194] and standard multiclass support vector machines (SVM) are used for comparison. The 2000 batch size are treated as 2000 features for the SVM.

Table 9 – Comparison with Documented Diagnostic Model

Model	**Validation Accuracy**	**Test Accuracy**
Proposed Trial 1	99.81%	91.73%
Proposed Trial 2	100%	90.98%
Proposed Trial 3	100%	86.47%
Proposed Trial 4	100%	90.23%
Proposed Trial 5	100%	90.98%
Proposed Average	**99.96%**	**90.08%**

Benchmark Trial 1	100%	88.72%
Benchmark Trial 2	86.28%	73.68%
Benchmark Trial 3	97.56%	84.21%
Benchmark Trial 4	100%	87.97%
Benchmark Trial 5	99.62%	88.72%
Deep Learning Benchmark Average	**96.69%**	**84.66%**
SVM Average	**78.38%**	**76.51%**

It can be observed that the proposed deep convolutional network with Bayesian optimization has outperformed one of the documented algorithms. The Bayesian optimization improves the prediction accuracy around 6%. The proposed algorithm can be implemented in various diagnostic applications for rotating machineries.

2.6.5 Conclusions

The detection of early fault is a challenging topic in the diagnosis of bearings. Documented deep learning based diagnostic method uses the unprocessed signal to perform diagnosis which could lead to inaccurate result because of the intrinsic degradation nature of bearings. The proposed deep convolution neural network uses the signal processed by

the CEEMD to extract the degradation pattern of bearings. By adding the bootstrap resampling algorithm, the CEEMD processed signal shows a clear trend of defect growth. The experimentally based defect size estimation equation is utilized to calculate relative defect size. The bearing degradation stage is classified based on the user requirement. The labelled segmented post-processed vibration data is input to the deep convolutional neural network for data classification. The Bayesian optimization algorithm is implemented to improve the model accuracy. By comparing with an existing model, the proposed algorithm shows a 6% improvement against a previously developed diagnostic model using deep learning algorithm in testing data accuracy and 18% improvement in accuracy versus the SVM algorithm. The optimized CEEMD algorithm automatically calculate the parameter of the added white noise. In addition, the Bayesian algorithm optimizes the deep learning network parameters without knowledge of tuning the sophisticated network. The proposed algorithm can be implemented in various health monitoring applications of rotating machineries. One thing needs to be noted is that the Bayesian optimization algorithm could possibly lead to local extrema. Therefore, the selection of parameter range needs to be thoroughly examined. In the future study, increasing the data range should be performed to check if an optimal solution set exists outside the range specified in this research work. In addition, a further refinement of the degradation stage should be tested. If the degradation stage is further divided and predicted, the diagnostic model can possibly be implemented in predicting defect size, which remains a great challenge for the condition-based maintenance.

CHAPTER 3. PROGNOSIS FOR ROLLING ELEMENT BEARING

The prognosis is implemented after the diagnosis to predict the RUL of the bearing or the degradation trend. The literature related to prognosis is not as comprehensive as diagnosis. The early research started around 1980s by Dyer and Steward. They used the kurtosis value as an indicator for bearing damage [195]. They concluded that the information provided by using the kurtosis permits a form of assessment of conditions, but it does not predict the failure time. Randall analysed the vibration spectra and reported that the early fault can be detected by the proposed method [196]. Statistical methods are generally used to categories the three stages of bearing life [197]. Documented research demonstrates that the bearings either fail early in the pre-failure stage or last for a long period of time and then start to degrade drastically. It is possible to predict the failure under known conditions for a specific type of bearings. However, in the actual application, unknown running conditions such as machine shutdown, loss of lubrication, harsh running conditions always occur. Therefore, intelligent prognosis comes into play. Prognosis is based on critical features extracted from the diagnosis. The general used features are the root-mean-square (RMS), kurtosis, crest factor, skewness of the signal. The trend of degradation are generally fitted into a physical model such as the Paris' law or a statistical model such as the auto-regressive and moving-average (ARMA) model [198, 199]. These models have deficiency of adaptation to various machine running conditions. Therefore, adaptation and modification of the models are implemented to compensate for the incapability of the general models. Li proposed using the recursive-least-square (RLS) to

update the coefficients of the Paris' law model [7]. The model overcomes the adaptation issue of the traditional model but has issues of the rate of convergence. Shao and Nezu implement a forward prediction neural network model to predict the degradation trend of the bearings [200]. This method is good for short time forecasting in the late stage of the bearing degradation. The current research can be concluded into three categories: 1. Model-based prognosis. 2. Experience-based prognosis. 3. Data-drive prognosis. The third category is the most popular in the recent years because of the progress in machine learning and artificial intelligence. The support vector machine, artificial neural network and self-organizing maps are the most widely used. The documented methods, in general, require offline calculation because of the computation complex. The future work will involve implementation of newly developed algorithm in an online manner.

3.1 ADAPTIVE PROGNOSIS OF BEARING DEGRADATION BASED ON WAVELET DECOMPOSITION ASSISTED ARMA MODEL

3.1.1 Introduction

Bearing failure is a very common issue exists in rotating machinery such as automotive axles, helicopter rotating shafts, gear reduction boxes. Industries suffer from the unpredicted bearing failures resulting in unscheduled down-time, catastrophic incidents and even personnel casualties. Therefore, a robust and reliable method to predict bearing failure is necessary to reduce the potential cost brought by the failure of the bearings. Previous research of the prediction of bearing life can be divided into three categories. The Lundberg and Palmgren model was introduced based on stresses in combination with probability [201]. However, this model under-estimates the fatigue lives of the bearings.

The second category uses the Paris's Law for defect growth as the basis for bearing life prediction [7]. The Paris's Law, in general, requires measurement of the defect size, which is hard to implement in certain circumstance. The last category involves artificial intelligent methods such as Neural Network to predict the remaining life of the bearing [200]. The training of Neural Network normally requires a significant amount of data to generalize the model. However, the availability of the experimental data is relative limited.

To address these issues, a prognostic model is developed using wavelet decomposition to remove the noise in the measured vibrational signal of the bearing with adaptive auto-regressive and moving-average (ARMA) model. The adaptation utilized the Recursive Least-Square (RLS) algorithm that updates the coefficients of the ARMA model based on the newly available signal during the online monitoring to enhance prediction accuracy. In addition, the implementation of RLS does not require any knowledge of the coefficients of the ARMA model. This advantage reduces the complexity of the computation of the ARMA model coefficients in comparison with the generally used method, Yule-Walker equation. The Yule-Walker equation requires matrix inversion and certain amount of measured data to calculate the ARMA coefficients, which may increase the computation complexity. We tested the model using the experimental data of vibration signal of the bearing. The purposed model is shown in Figure 65.

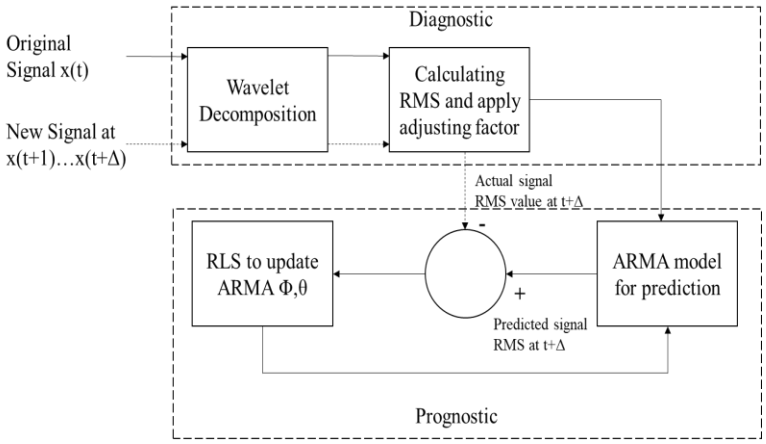


Figure 65 – Wavelet Decomposition ARMA+RLS Model

The original signal was processed using the wavelet decomposition de-noising and the RMS value was calculated using Equation (128) for every 1280 data point that is equivalent to 5 seconds and 125 revolutions of the bearing. The coefficients of the ARMA model is assigned randomly to initiate the model. When new signal is available later, the new RMS value is calculated and compared with the prediction from the ARMA model. The RLS algorithm calculates the updated ARMA coefficients based on a least square criterion and updates the ARMA coefficients to achieve the adaptation in the model.

$$x_{rms} = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + \cdots + x_n^2\right)} \tag{128}$$

3.1.2 Wavelet Decomposition

The wavelet decomposition algorithm [15] has shown successful implementation in bearing prognostics. Qiu implemented a wavelet based de-noising mechanism to enhance the signal to noise ratio and extract important feature in the simulated vibrational signal [202]. He developed a super-wavelet method to enhance feature extraction on

induction motor [203]. Generally, the wavelet decomposition procedure consists of the following steps: For a series of data, x(n), with sample rate f_s, the wavelet decomposition divides the data into an approximated signal (0- $f_s/4$) and detailed signal ($f_s/4$- $f_s/2$). The next decomposition divides the approximated signal into a new approximated signal (0- $f_s/8$) and detailed signal ($f_s/8$- $f_s/4$). The process can be implemented repetitively to decompose the signal into multiple layers. The general equation is shown below:

$$a_n^j = \sum_l h(l-2n)a_l^{j-1} \tag{129}$$

$$d_n^j = \sum_l g(l-2n)a_l^{j-1} \tag{130}$$

where a_n^j is the j-th layer approximated signal, and d_n^j is the j-th layer detailed signal, h is the low pass filter and g is the high pass filter. The schematics of the decomposition is shown in Figure 66.

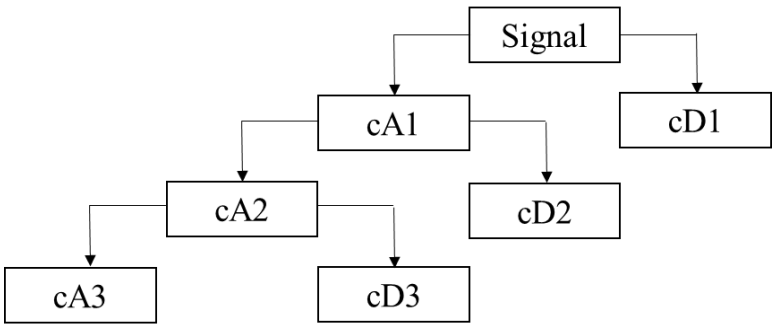


Figure 66 – Wavelet Decomposition Schematics

3.1.3 ARMA Model

The ARMA model is widely used to predict time series across different fields. The ARMA model consists of two parts: auto-regressive (AR) and moving average (MA). A general representation of an ARMA(p,q) model is shown in Equation (131):

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{131}$$

where c is the ARMA model constant, ε_t is an independent identically distributed white noise sampled from a normal distribution with a mean value of zero, x_t is the RMS value of the vibrational signal, φ_i and θ_i are the coefficients for the AR and MA model respectively [204]. A widely accepted method to determine the ARMA model parameter p and q is the Akaike Information Criterion (AIC) [205]. The AIC states the ARMA (p,q) model having the lowest value of the information criteria value is the most efficient model to represent a time series. In this section, we implemented an ARMA (2,2) as the initial model to predict the RMS value of the bearing vibrational signal.

3.1.4 RLS Algorithm

The RLS algorithm in this model is initialized using the inverse correlation matrix **P**, which is a diagonal matrix with large value of the diagonal terms [206]. For the ARMA (2,2) model with a constant c, **P** is a 5×5 matrix with all the diagonal terms equal to 100. The weight-vector $\hat{\boldsymbol{w}}$ was initially set to random value because no previous knowledge is needed to determine the ARMA coefficients with the RLS algorithm implemented [7]. For each RMS value, the RLS algorithm calculates the gain vector **k** as shown in Equation (132):

$$\mathbf{k}(n) = \frac{\lambda^{-1}\mathbf{P}(n-1)\mathbf{u}(n)}{1+\lambda^{-1}\mathbf{u}^{H}(n)\mathbf{P}(n-1)\mathbf{u}(n)} \tag{132}$$

where λ is the forgetting factor with a normal range from 0.8 to 1, **u** is the input vector with a size equals to number of coefficients needs to be updated. In this case, **u** consists of the ARMA constant, the previous two RMS values and the previous two white noise values. The error term is presented using equation (133):

$$\xi(n) = d(n) - \hat{\mathbf{w}}^{H}(n-1)\mathbf{u}(n) \tag{133}$$

with *d* represents the actual RMS value measure from future input to the adaptive model, $\hat{\mathbf{w}}^{H}(n-1)\mathbf{u}(n)$ represents the predicted value of the ARMA model using the previous ARMA coefficients. The RLS algorithm reduces the error between the actual value and the predicted value to minimum by changing the weight-vector ***w***. ***w*** is updated accordingly based on equation (134) and the inverse correlation matrix updates itself according to equation (135):

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\xi^{*}(n) \tag{134}$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{u}^{H}(n)\mathbf{P}(n-1) \tag{135}$$

a significant advantage of this algorithm, as mentioned before, is no previous knowledge of the system is needed, which reduces the effort of modeling complex systems and enables online monitoring processes. One thing to be noted is that the RLS algorithm does not guarantee to converge to the true value of the system. Instead, it converges to a combination of values to minimize the least square error.

3.1.5 Experimental Setup

Experimental vibration data of the bearing is obtained using the data published by Nectoux *et al.* [207]. Figure 67 shows the experimental setup. The vibration signal was collected using a DYTRAN 3035B accelerometer in the horizontal and vertical direction with a 25.6kHz sampling frequency and each 2560 data point was recorded every 10 seconds. The bearing was running at 1500RPM with 5000N load.

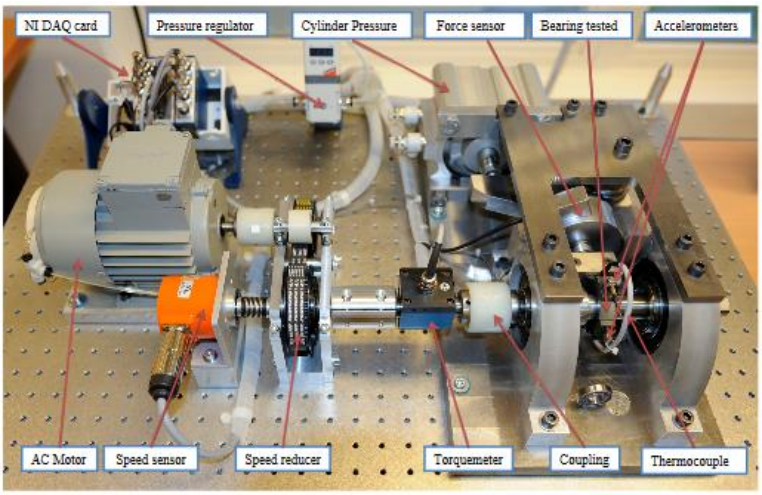


Figure 67 – Experimental Setup for Monitoring of Bearing Vibration [207]

3.1.6 Result

Figure 68 shows experimental vibrational signal of the bearing. To reduce the noise in the signal, a level 5 wavelet decomposition method is implemented. The reconstructed signal in comparison with the original signal is shown in Figure 69. The error between the original signal and reconstruction signal is relatively small, which indicates minimal distortion of the original signal. The signal to noise ratio (SNR) is calculated to be -1.62dB.

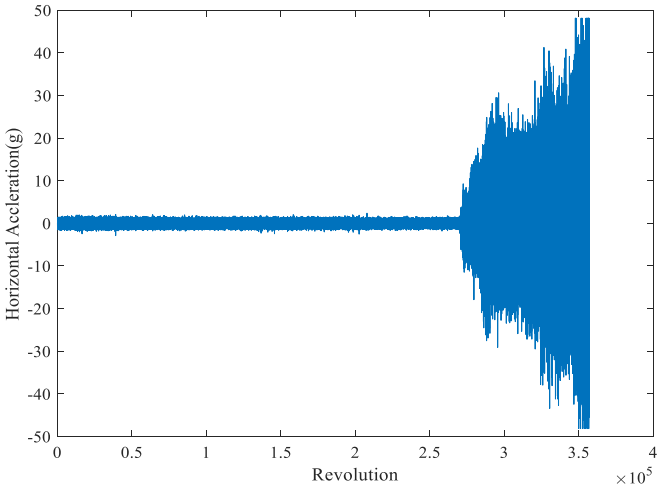


Figure 68 – Unprocessed Vibrational Signal of the Bearing

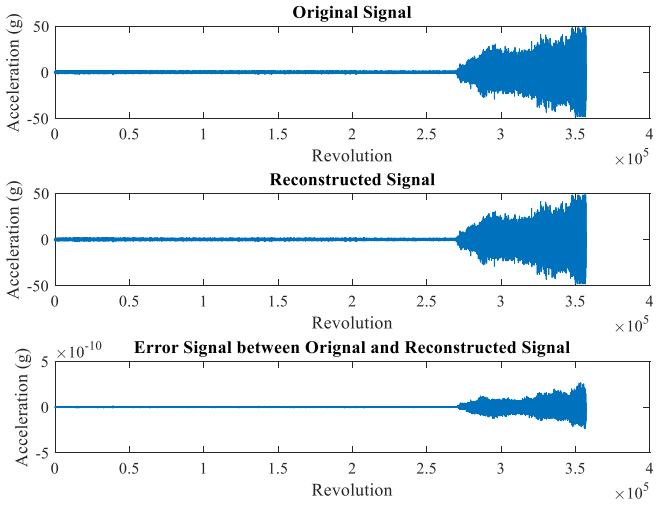


Figure 69 – Reconstruction of Signal after Wavelet Decomposition

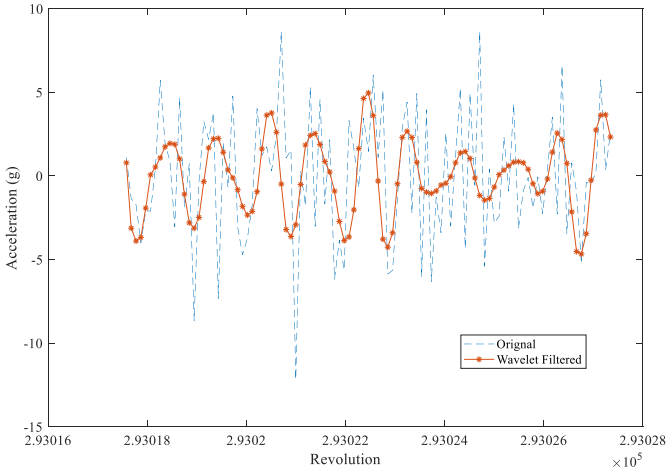


Figure 70 – Decomposed Signal

Figure 70 shows a section of the wavelet filtered reconstructed signal and original signal. The signal after processing maintains its original degradation trend after the filtering. The RMS value was calculated using every 1280 data point. The RMS value after filtering is shown in Figure 71.

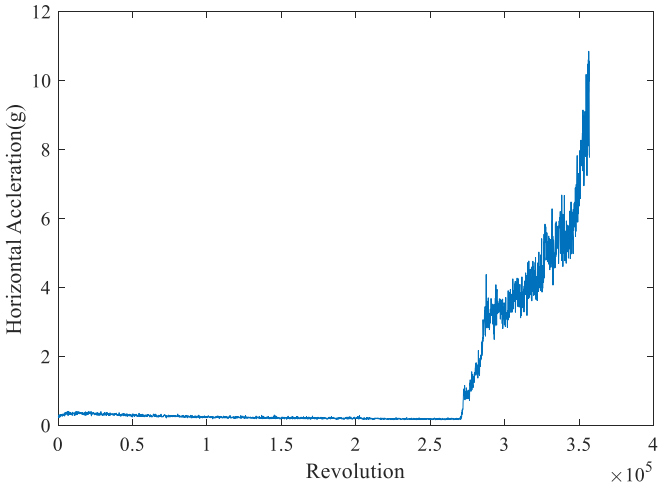


Figure 71 – RMS Value of Vibration Signal

The predicted signal versus the experimental signal is shown in Figure 72. The errors between the measured and predicted signal are calculated and shown in Figure 73 and Table 10.

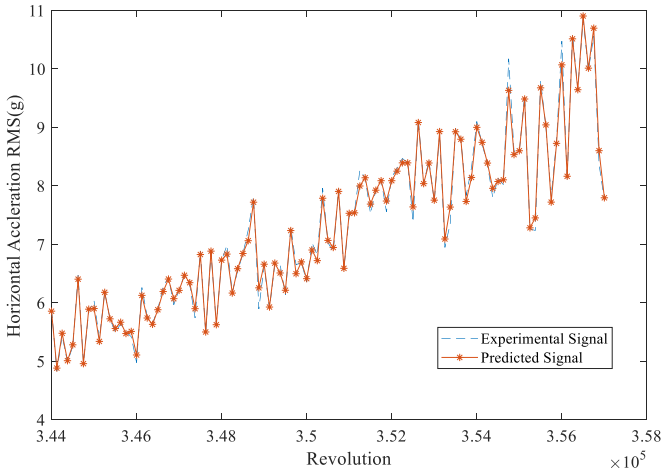


Figure 72 – Experimental Signal vesus Predicted Signal

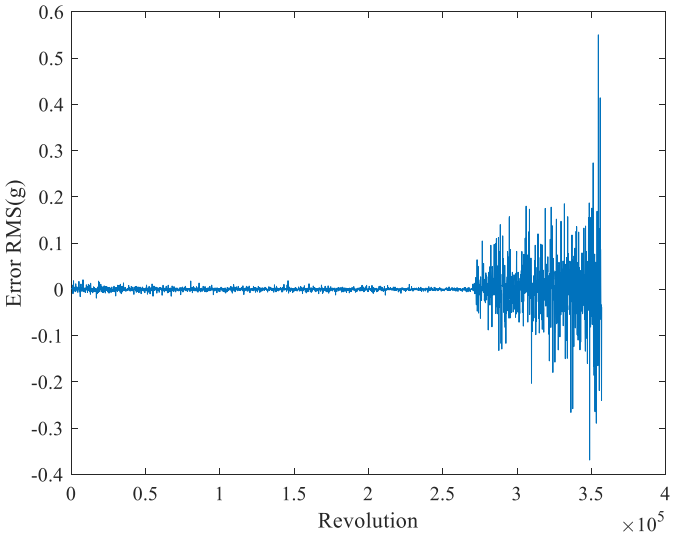


Figure 73 – Overall Error of RMS Value

Figure 73 shows the overall error of prediction during the bearing degradation process. The largest error occurs at the bearing crack propagation stage at the end of the running cycle. The model prediction shows relatively good predicting result as shown in Figure 72. In addition, we investigated the effect of different values of forgetting factors and multiple iterations of training for the RLS algorithm. The result is shown in Table 10. *i* is the

iteration number for the RLS algorithm. The mean square error (MSE) between the predicted value and experimental value using a forgetting factor of 0.95 and iteration number of 10 is 0.0013. The mean average percentage error (MAPE) is 1.03% and the mean absolute error (MAE) is 0.56. The experimental value of the RMS is 10.17 when the maximum prediction error occurs. An error of 0.56 is 5.5% of the experimental RMS value. Therefore, the proposed model shows good prediction accuracy in the bearing degradation process. Furthermore, the error can be reduced by adjusting the forgetting factor and the iteration number. The effect of the forgetting factor will be investigated in [208].

Table 10 – Error with Various Forgetting Factor Values

	MSE	**MAPE**	**MAE**
Forgetting Factor=0.80, i=10	7.55×10^{-9}	0.0012	0.0029
Forgetting Factor=0.90, i=10	3.19×10^{-5}	0.13	0.13
Forgetting Factor=0.95, i=10	0.0013	1.03	0.56
Forgetting Factor=0.99, i=10	0.021	4.43	1.60
Forgetting Factor=0.90, i=3	0.0076	2.63	1.31
Forgetting Factor=0.90, i=30	6.41×10^{-12}	2.54×10^{-5}	8.98×10^{-5}
Forgetting Factor=0.99, i=50	0.0012	0.98	0.55

3.1.7 Conclusions

This section purposes a wavelet assisted adaptive bearing prognostic model using ARMA model with RLS. The ease of implementation and minimal knowledge of the modeling system enable this model to be applied in online bearing monitoring. The wavelet decomposition is used for signal denoising to enhance the signal to noise ratio. The ARMA model in combination with the RLS can predict the RMS value of the bearing vibration. The adaptive model overcomes the inflexibility of the deterministic models and stochastic models by constantly adjusting itself based on a least mean square criterion. The error between the prediction and experimental result is calculated to validate the model. The model shows good prediction result with an MSE of 0.0013, MAPE of 1.03% and maximum prediction error of 0.56 in the RMS value. In addition, the error can be reduced by increasing the iteration number. The effect of the forgetting factor will be investigated in the future study. We believe that there is an optimal value of the forgetting factor, which can reduce error while using the minimal iteration number.

3.2 PROGNOSIS OF BEARING DEGRADATION USING GRADIENT VARIABLE FORGETTING FACTOR RLS COMBINED WITH TIME SERIES MODEL

3.2.1 Introduction

Various mechanical systems such as gearboxes, helicopter rotors, and spindle assemblies of CNC machines rely on the running condition of the bearings. Even though the components of the bearings are made of materials with superior mechanical strength and fatigue life, limitations of the capabilities of the manufacturing process along with the

defects of material lead to the formation of microcracks on the surfaces and sub-surfaces of the elements of bearings. Over time, the cracks will propagate because of periodic loadings and subsurface plastic flow. In addition, lack of lubrication, contamination, misalignment, corrosion, and improper loads cause premature failure of bearings. Because of its highly unpredictable failure time and mode, bearing failure is a major issue in rotating machineries. Therefore, companies are devoting significant effort to estimating the severity of bearing damage and predicting the failure time and degradation trend of bearings by implementing bearing diagnostic and prognostic schematics. Traditionally, the severity of bearing damage, such as the size of the defect, is measured by placing bearings under certain running conditions on a testing rig, disassembling the bearings after a period of time and measuring with an optical device [7]. However, the measurement of defect size is time-consuming. Thus, unobtrusive signal acquisition methods that use accelerometers or acoustic emission sensors are preferred. Because the time-domain acceleration data clearly reflect the degradation trend and are easily obtainable, the signal acquired by accelerometers is used in this section to track the degradation process.

Bearing diagnosis can be classified into three major groups: time-domain, frequency-domain, and time-frequency-domain analysis. Time-domain methods use an acquired signal from sensors placed on a component of the machine to extract critical features for the diagnosis of the bearing condition. Some desired features of the diagnosis in the time-domain analysis are shock pulse counting, root mean square (RMS), peak values, crest factors, kurtosis, short-time energy, and short-time zero-crossing rate. Early research has demonstrated that time-domain data indicate certain patterns of the types of defects [7, 209-212]. Using Fourier transform, researchers find that the frequency-domain data reveal the

critical frequencies of vibration signals. Some widely used methods include bi-coherence analysis, cepstrum analysis, and the high-frequency response technique (HFRT) [7], which has been widely implemented. Using the HFRT, one study found that the impact generated by the defect site in the bearing elements normally excites resonance in other components in the system [48]. The high-frequency components directly reflect the damage levels of bearing during its service life.

The fundamental characteristic frequencies of the rolling elements of bearings are often buried during the signal monitoring because of the existence of noise and resonance. Therefore, researchers implemented different signal processing techniques to denoise the undesirable signals and locate important information in the system. For instance, Shiroishi *et al*. [141] utilized an adaptive line enhancer that increases the signal-to-noise ratio to detect small defects; Zhou *et al*. [213] implemented the Wiener filter to extract the bearing fault signature. Tian implemented spectral kurtosis and cross-correlation to detect incipient faults and location of faults without reference data [214]. Tian also implemented simulated annealing to optimize the spectral kurtosis to locate optimum frequency band for diagnosis [215]. The purpose of bearing prognostics is to predict the fault of the bearing before its catastrophic failure [137]. The most critical factors for optimizing the bearing maintenance schedule and further reducing cost are the accuracy and reliability of the prognostic methodology. Early researchers developed deterministic models from the physical understanding of the defect-propagation processes based on fracture mechanics [216, 217]. However, the propagation of defect is stochastic because most manufacturing processes induce phase transformation and grain structure change [218].

One of the most common prognoses is the prediction of the remaining useful life (RUL) of bearings. For instance, Qiu *et al*. [16] compared the performance of wavelet decomposition-based de-noising and wavelet filter-based de-noising methods in bearing prognosis. Caesarendra *et al*. [219] applied relevance vector machine and auto-regressive and moving-average (ARMA) model in simulated bearing data. All these methods are relatively complex and require a significant amount of computation, which results in issues for online application. To define the bearing failure point, researchers choose a specific size of defect to quantify damage of bearings [7, 141]. Shiroishi *et al*. [141] proposed the failure size as 0.01 in^2. However, this type of defect size should not be generalized to all types of bearings because the size of the defect is proportional to the size of the bearing elements. Therefore, defining a damage severity level for a specific type of bearings is superior to using the pre-defined defect size as the criteria of failure.

This study uses an innovative combination of the ARMA model with the variable forgetting factor recursive least-square (VFF-RLS) algorithm to achieve prognosis of the vibrational signal of the bearings. In addition, the model overcomes the computational complexity and adapts to the stochastic nature of the fatigue behavior of bearings. The ARMA model is implemented instead of the ARIMA model because of the simplicity and commonality of the model in online application. The gradient based variable forgetting factor improves the convergence speed of the RLS algorithm and yields smaller error with fewer iterations of training. To prepare the signal for prognosis, a Butterworth band-pass filter and local regression smoothing filter is used along with the RMS of the signal to track the behavior of the bearing in the time-domain. The prognostic model implements the gradient based VFF-RLS algorithm to update the coefficients of the ARMA model, which

predicts the degradation trend during the online monitoring process. Adaptation is needed during the online monitoring because the bearing degradation process does not exhibit a consistent increasing trend, and the process itself is highly stochastic. Even though vibrational signal in the time-domain is non-stationary, the in-process adaptation of the VFF-RLS algorithm tracks the changes of vibrational signal and reduces errors between the predicted and experimental values accordingly.

The rest of the section is organized as follows: Section 2 describes the models used in bearing prognosis and introduces the ARMA+VFF-RLS model used in this section. Section 3 includes testing the applicability of the ARMA+VFF-RLS model in online monitoring of the vibration of the bearings. To further validate the model's applicability in various conditions, the ARMA+VFF-RLS model was tested using simulated data in Section 4. Section 5 shows the improvement of the multi-iteration training on individual data points and compares the traditional RLS with the VFF-RLS. In addition, the ARIMA model is used as a benchmark to prove the improvement of the ARMA+VFF-RLS algorithm. Section 6 concludes the section with comments and suggestions for future work.

3.2.2 ARMA+VFF-RLS Model

3.2.2.1 Deterministic Crack Propagation Model

A deterministic bearing diagnostic model is developed based on fracture mechanics and failure analysis for high-cycle fatigue. Various experimental data show that crack growth of bearing element under running conditions depends on a variety of factors such as material properties, loading profile, manufacturing processes, size of bearing elements

and other environmental effects. The most popular empirical model is Paris' law [220] shown as (136),

$$\frac{da}{dN} = C_0(\Delta K)^n \tag{136}$$

where $\frac{da}{dN}$ is the rate of the crack extension, instantaneous half crack length is denoted by *a*, the number of running cycles is represented by *N*, C_0 and *n* are constants related to the material properties and running conditions, and ΔK is the stress intensity factor range. Over the last decade, different researchers modified (136) to capture the failure mechanisms of bearings. For instance, Li *et al*. [221] applied a log-normal random variable to the original Paris' law to capture the stochastic nature of the defect propagations of bearings. This modified equation follows the general trend of the failure processes of bearings under stationary running conditions. However, other experimental data indicate that the failure of the bearing does not always follow Paris' law closely [207]. The measured vibration signals generally remain constant for a long period of time before sudden failure occurs. In addition, under certain circumstances, such as sharp edge rounding of the defects and reduction of the rotation speed, the measured vibrational signal decreases as shown in Figure 74. Because of the decreased vibration signal, the initiation of defect is extremely difficult to distinguish from the vibration signal. The deterministic models cannot resolve the two issues described. Therefore, implementing an adaptive model to capture the changes of the vibrational signal of bearings is beneficial.

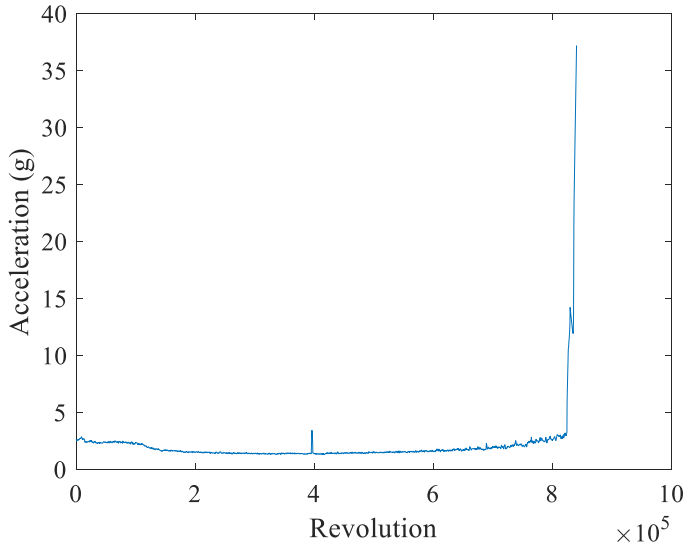


Figure 74 – RMS value of Vibrational Signal

3.2.2.2 Adaptive Prognostic Model

The prognostic model implemented reduces the complexity of computation compared with other models, such as the ARIMA model, in the online monitoring application. The proposed model is shown in Figure 75.

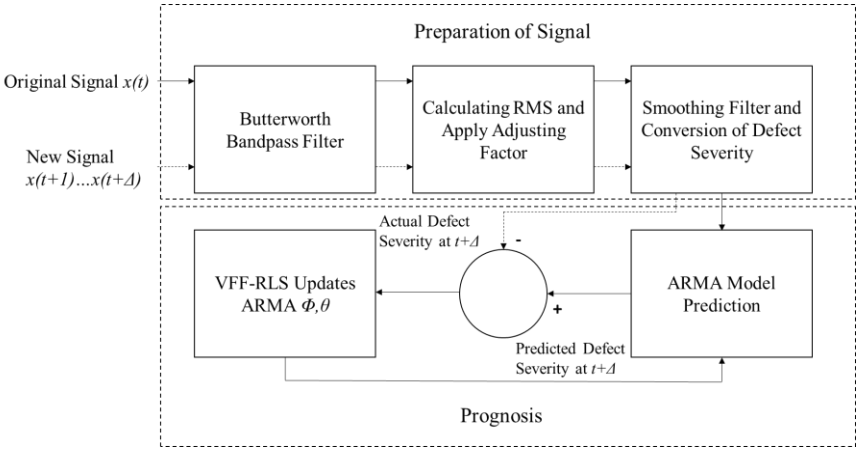


Figure 75 – Adaptive Bearing Diagnostic and Prognostic Model

The model first takes in the signal from *x(0)* to *x(t)* in terms of output voltage of the accelerometer and passes the signal to the bandpass filter to eliminate high-frequency noise. The future signals *x(t+1)…x(t+Δ)* pass the same filter and are used to generate error terms for the RLS algorithm. After the signal is filtered, the RMS value of the vibration signal is calculated based on (137),

$$x_{rms} = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + \cdots + x_n^2\right)} \tag{137}$$

where n is the number of data points used in the calculation, and x_1, x_2 … x_n are the vibrational signals. The selection of n will affect the RMS of the vibrational signals. After the RMS value is calculated, a robust local regression smoothing filter is implemented to smooth out the leftover narrow-band oscillation in the RMS values. The regression smoothing filter computes regression weights for individual data points and uses linear least-squares regression to obtain the smoothed data. The magnitude of the RMS decreases after the filters are applied. Therefore, after completion of the filtering, a correction factor is applied to the filtered RMS to compensate for the decreased magnitude. The last step of

the signal preparation is to convert the RMS value to either defect size or defect severity. Li [7] has identified that the value of RMS, kurtosis, and peak force of the vibration signal within a certain frequency range is related to the defect size. The relationship between the RMS and the defect size shows either a linear relationship with fitness value of $R^2=0.850$ or an exponential relationship with $R^2=0.843$. Similarly, a power law and a second order polynomial regression can be implemented between the kurtosis and the defect size using the measurement from Li [7] and Shiroishi *et al*. [141]. Because of the similarity in the experimental setup and bearing type with [7], the relationship between the RMS value and defect size in this section can be described using (138),

$$D = a \times RMS + b \tag{138}$$

where D stands for the defect size (area) or defect severity, a is a coefficient related to a specific bearing running condition, and b is a constant with respect to the bearing type and the running condition. In the application of prognosis, an experiment which determines the values of coefficient a and constant b is necessary to establish the relationship between the severity of defect and RMS of the signal.

The adaptive part of the model uses the ARMA model to predict the signal at time $t+\Delta$. Once the newly acquired signal is available, the error between the experimental and predicted signal is obtained. The error is passed to the VFF-RLS algorithm to correct the coefficients of the ARMA model for future prediction. The adaptive part of the model is implemented recursively so that the coefficients of the ARMA model are updated constantly during the online monitoring. The model overcomes the restrictions and limitations of a deterministic model and represents a robust and innovative method for

bearing prognosis. The following section describes the ARMA model and VFF-RLS algorithm used in the prognostic model.

The ARMA model includes two parts: auto-regressive (AR) and moving-average (MA). An ARMA(p,q) model can be described using (139),

$$x_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i x_{t-i} - \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{139}$$

where c is a constant and ε_t is an independent identically distributed random variable sampled from a normal distribution with zero mean value, $\varepsilon_t \sim N(0,\sigma^2)$. The variable σ^2 is the variance, ε_t is normally treated as a white noise with a variance of σ^2, x_t stands for the vibrational data in the time domain, and φ_i and θ_i are the coefficients of the AR and MA model respectively [222]. A general way of determining the ARMA model parameters p and q is the Akaike Information Criterion (AIC). The ARMA (p,q) model with the lowest AIC value is the most efficient model to describe a process.

The RLS algorithm can be initialized by setting $\boldsymbol{P}(0)=\delta^{-1}\boldsymbol{I}$, where $\boldsymbol{P}$ is the inverse correlation matrix, δ is a small number generally used as a scaling factor, and $\boldsymbol{I}$ is an identity matrix [223]. The size of $\boldsymbol{I}$ corresponds to the number of coefficients that the RLS algorithm updates. For instance, an ARMA (2,2) model described by (139) has five coefficients: one constant, two coefficients from the AR model, and two coefficients from the MA model. In this case, $\boldsymbol{I}$ represents a five by five identity matrix. The tap-weight vector, $\hat{\boldsymbol{w}}(0)$, was set to zero because the initial value does not affect the convergence of the RLS algorithm [7]. The length of the tap-weight vector equals the number of coefficients of the ARMA

model. For an ARMA (2,2) model with five coefficients, the tap-weight vector $\hat{\boldsymbol{w}}(0)$ is a five by one vector. Then, for each instant of time, *n=1,2,3,…,* the algorithm calculates the following entities,

$$\mathbf{k}(\boldsymbol{n}) = \frac{\lambda^{-1}\mathbf{P}(\boldsymbol{n}-1)\mathbf{u}(\boldsymbol{n})}{1+\lambda^{-1}\mathbf{u}^{\boldsymbol{H}}(\boldsymbol{n})\mathbf{P}(\boldsymbol{n}-1)\mathbf{u}(\boldsymbol{n})} \tag{140}$$

where $\mathbf{k}$ is the gain vector, and λ is the forgetting factor. When λ is close to one, the RLS algorithm converges to a steady state error in a slow manner while yielding a small error, and for λ close to zero, the algorithm converges to a steady state error in a relatively fast manner while yielding a large error. The value of λ normally varies between 0.8 and 1. The input vector $\mathbf{u}$ has a dimensional size equal to the number of coefficients that need to be updated. Using the ARMA (2,2) example, $\mathbf{u}$ is a one by five vector. The error ξ between the experimental and predicted data is shown in (141),

$$\xi(n) = d(n) - \hat{\mathbf{w}}^{H}(n-1)\mathbf{u}(n) \tag{141}$$

where ***d*** is the experimental value, $\hat{\mathbf{w}}^{H}(n-1)\mathbf{u}(n)$ represents the predicted value at time *n*, and ξ^{*} is the transpose of ξ. The error term ξ is supposed to decrease to zero as the algorithm updates the coefficients through the tap-weights vector of the model.

$$\hat{\mathbf{w}}(n) = \hat{\mathbf{w}}(n-1) + \mathbf{k}(n)\xi^{*}(n) \tag{142}$$

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{u}^{H}(n)\mathbf{P}(n-1) \tag{143}$$

One advantage of this algorithm mentioned by Li *et. al* [7] is that the initial values of the coefficients that need to be updated do not affect the error convergence. In comparison with other algorithms, such as Akaike Information Criterion (LMS), the rate of convergence of RLS is an order of magnitude higher since the RLS algorithm does not require as many iterations to update the tap-weight vectors as does the LMS. This feature greatly improves its applicability in online bearing monitoring processes [223].

The gradient based variable forgetting factor algorithm improves the RLS algorithm convergence speed by changing the forgetting factor λ in (140). As demonstrated by So *et al.*, this algorithm overcomes the deficiency of the traditional RLS algorithm in time-varying models [224]. The gradient of the forgetting factor is obtained by deriving the dynamic equation of the mean square error. The algorithm adjusts the forgetting factor to minimize the mean square error and updates the forgetting factor recursively as shown in (144) [224],

$$\mathbf{P}(n) = \lambda^{-1}\mathbf{P}(n-1) - \lambda^{-1}\mathbf{k}(n)\mathbf{u}^{H}(n)\mathbf{P}(n-1) \tag{144}$$

where $\nabla\lambda(n)$ is the gradient of the forgetting factor with respect to the mean square error, and μ is the step size in the gradient-descent algorithm. The recursive equation has an upper limit $\lambda+$ and a lower limit $\lambda-$ to avoid the divergence of this algorithm. The gradient of the mean square error σ_e^2 with respect to the forgetting factor is shown in (145) [224],

$$\frac{\partial \sigma_e^2(n+1)}{\partial \lambda} = \lambda_0 \frac{\partial \sigma_e^2(n)}{\partial \lambda} + \frac{\partial C_1}{\partial \lambda}\sigma_e^2(n) + \frac{\partial C_2}{\partial \lambda}\sigma_\eta^2 \tag{145}$$

where λ_0 is the initial value of the forgetting factor, and σ_η^2 is the variance of the error between the experimental value and predicted value. With assumptions and simplifications, $\frac{\partial C_1}{\partial \lambda}$ and $\frac{\partial C_2}{\partial \lambda}$ can be simplified as (146) and (147) [224],

$$\frac{\partial C_1}{\partial \lambda} = 2 + \frac{2(N+2)(1-\lambda)[N\lambda^2 - (N+3)\lambda - 1]}{[N(1-\lambda)+2]^2} \tag{146}$$

$$\frac{\partial C_2}{\partial \lambda} = -2 - \frac{4(1-\lambda)[N\lambda^2 - (N+3)\lambda - 1]}{[N(1-\lambda)+2]^2} \tag{147}$$

where *N* is the length of the RLS filter. The choice of *N* depends on the signal to noise ratio. No specific guidance exists to select the optimized *N* number. For this ARMA(2,2) model, the filter length is set to five by trial and error. The VFF-RLS algorithm significantly reduces the number of iterations to minimize the mean square error as described in [224]. In the next section, the diagnostic and prognostic model is tested using an experimental bearing vibration data.

3.2.3 Experimental Data

To investigate the feasibility of the proposed approach for online monitoring of bearings, bearing data measured by Nectoux *et al.* [207] is used for examining the proposed algorithm. Figure 76 shows the raw signal in the time domain. The data show that the raw experimental signal is comparatively noisy.

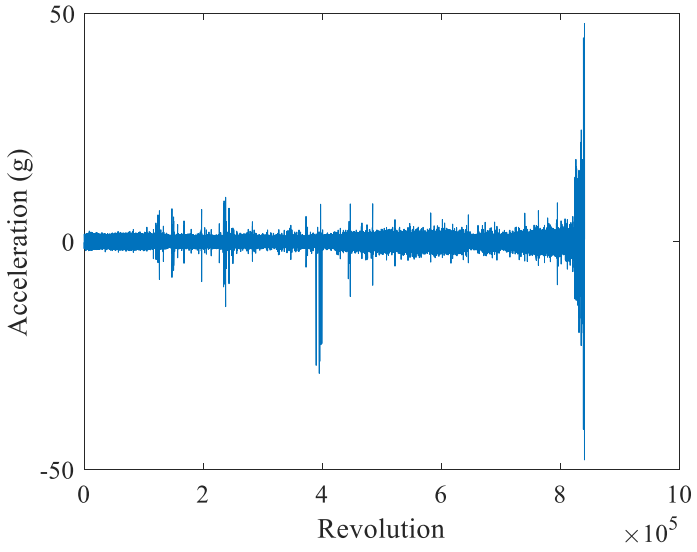


Figure 76 – Unprocessed Vibrational Signal

Because only using the VFF-RLS algorithm for noisy signals generates undesirable oscillations in predictions, bandpass and smoothing filters are implemented to exclude the noise within the data and facilitate the convergence speed of the VFF-RLS algorithm. The filtered signal using a fifth order Butterworth bandpass filter (1 kHz to 5 kHz) is shown in Figure 77. We picked the frequency range of the filter based on previous experiment conducted by Li [7], which indicates that the high frequency components reflect the defect signal clearer than low frequency components in the time domain.

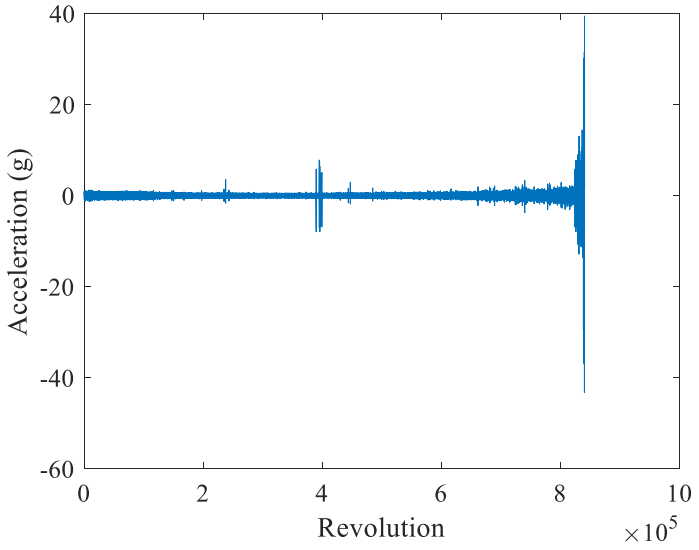


Figure 77 – Vibrational Signal Bandpassed from 1 kHz to 5 kHz

Since the magnitude of the signal is decreased by using the Butterworth filter, the signal requires rectification before it is used to calculate the RMS. Each value of the RMS is obtained using 2560 sample points. The resulting RMS after applying a correction factor is shown in Figure 78.

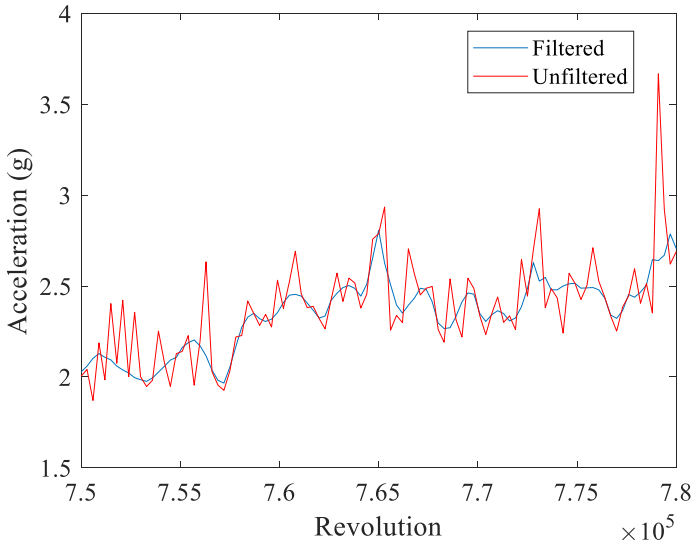


Figure 78 – Comparison of RMS Value with and without Filtering

The smoothing filter eliminates most of the narrow band noises, which reduces fluctuation in prediction. After the smoothing filter has been successfully implemented, the conversion equation between the RMS value and the defect severity is applied based on the user's requirement. The conversion between the RMS value and defect severity uses a similar structure as shown in (138). A defect severity level from one to ten is assigned based on the failure requirement in the experiment performed by Nectoux et. al. [207], with a severity of ten indicating a bearing failure. The experimental defect severity is shown in Figure 79.

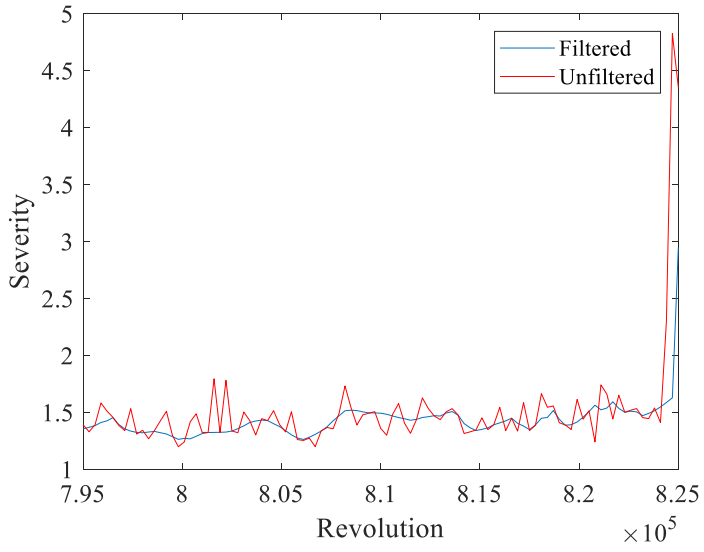


Figure 79 – Smoothed Experimental Defect Severity

To find the most efficient model to present the signal, the AIC is implemented to find the ARMA model structure. The AIC test result suggests an ARMA (2,2) model structure for this specific experiment. The VFF-RLS correlation matrix parameter δ is selected to be 0.01 and the forgetting factor λ is initially set to 0.9. The influence of the convergence of error by λ and δ is demonstrated by Zhang [225], which mentioned that a large forgetting factor reduces the steady state error while sacrificing the speed of convergence. The benefit of the VFF-RLS is that it overcomes the deficiency of the fixed value forgetting factor. Figure 80 and 81 show the predicted trend and modeling error during the prognostic process with δ=0.01. The spikes from Figure 80 of the prediction error indicate a sudden change in the experimental signal, which possibly indicates damage.

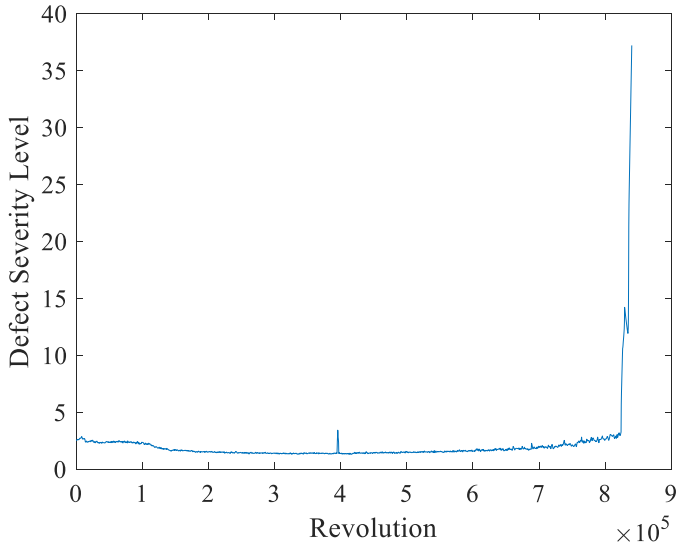


Figure 80 – ARMA+VFF-RLS Prediction with δ=0.01 and λ=0.9

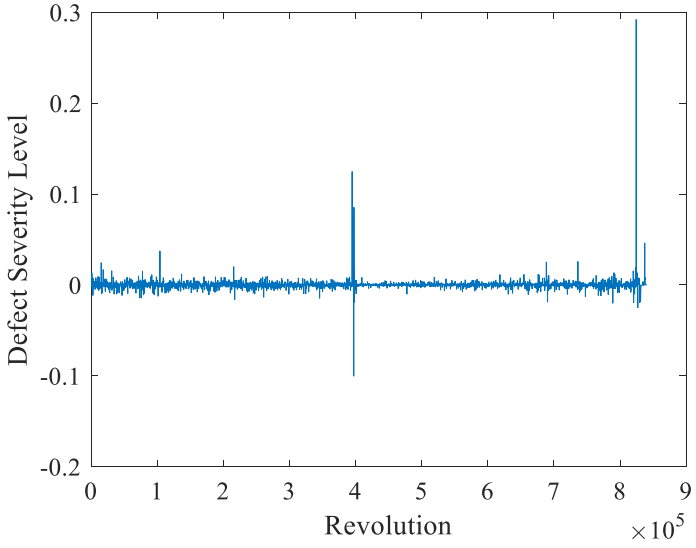


Figure 81 – ARMA+VFF-RLS Error with δ=0.01 and λ=0.9

By comparing Figure 77 and 81, we observed that the sudden increase of error around four hundred thousand cycles in Figure 81 is related to an unusual increase of the signal in Figure 77. Because the duration of the increased error is short, we attribute this change into possible operational error while conducting the experiment. Figure 82 and 83

are an enlarged portion of Figure 80 and 81 from eight hundred twenty thousand cycles to eight hundred forty thousand cycles. Although large fluctuations are observed during the updates of the coefficients of the ARMA model, the relatively fast convergence feature of the VFF-RLS algorithm efficiently reduces the error to zero as shown in Figure 83 around eight hundred twenty-five thousand cycles.

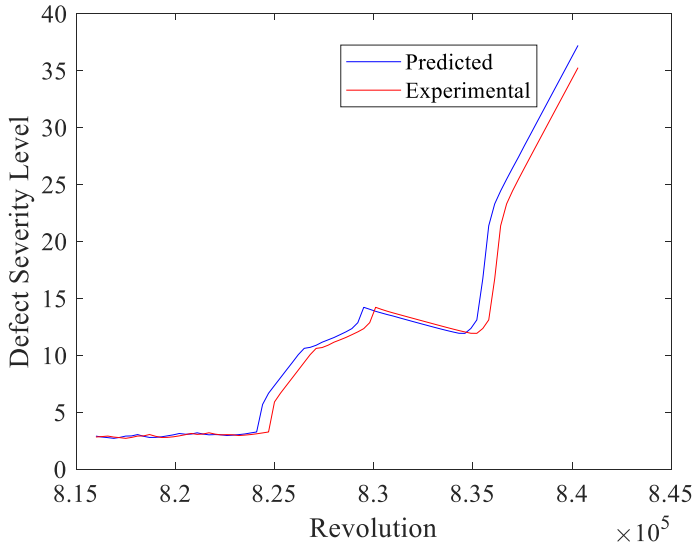


Figure 82 – Zoomed Interval of Prediction

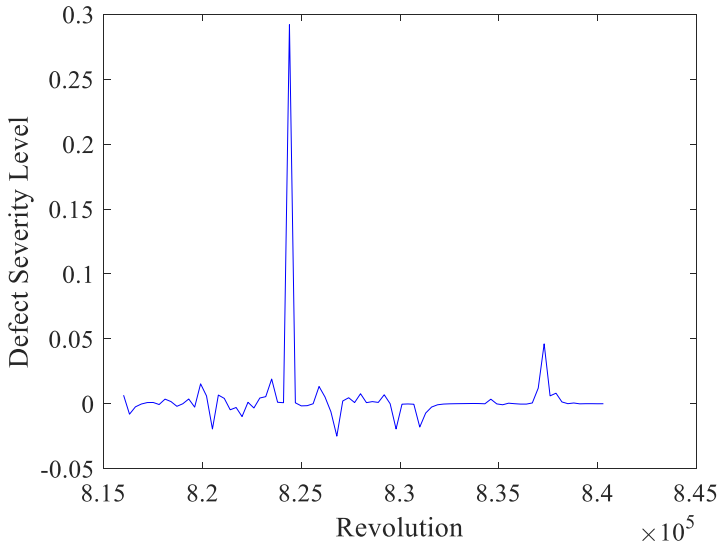


Figure 83 – Zoomed Interval of Prediction Error

3.2.4 Simulated Data

The model was also tested on simulated bearing data created using an exponential function with sinusoidal functions of various frequencies. The simulated bearing signal can be seen in Figure 84.

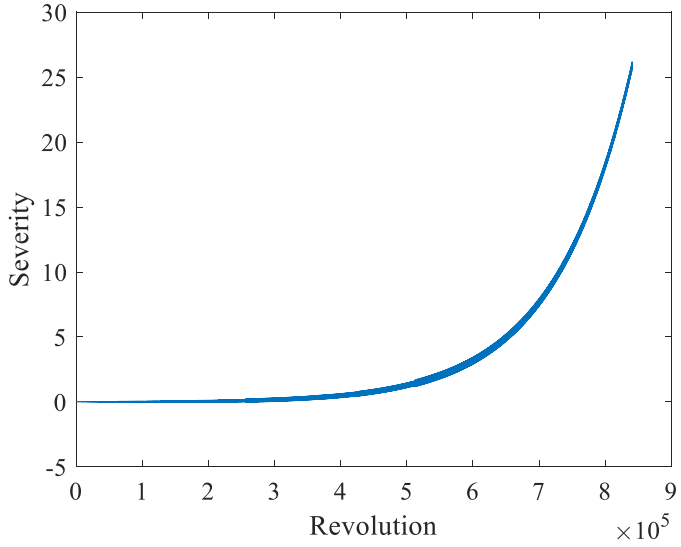


Figure 84 – Simulated Data

Using the same diagnostic model as shown in Figure 75, the smoothed RMS is shown in Figure 85. The discontinuity of the RMS value in Figure 85 is because of the window size of calculating the RMS. A smaller window could be used to reduce the discontinuity shown in Figure 84.

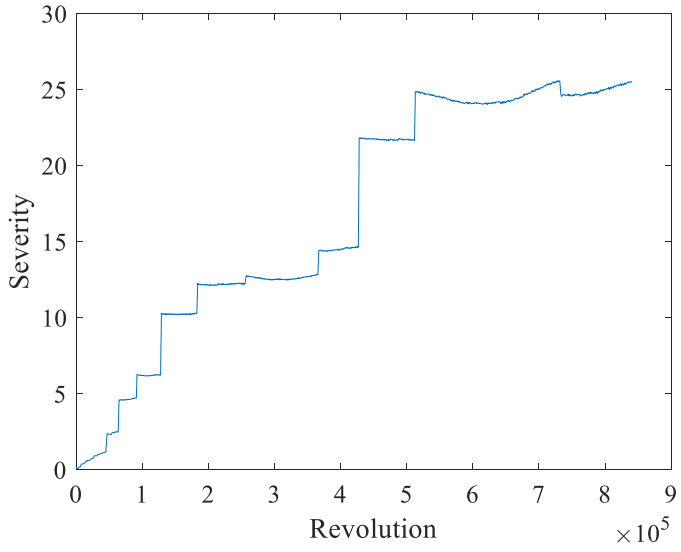


Figure 85 – RMS Value for Simulated Data

In addition, the prediction error of the ARMA+VFF-RLS is shown in Figure 86. Both the simulated and experimental data demonstrate the successful prediction of the bearing failure using the ARMA+VFF-RLS model in different patterns of degradation.

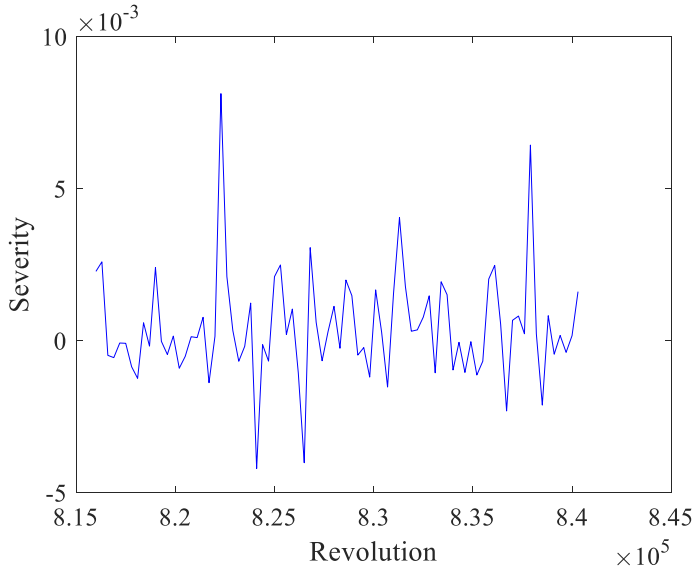


Figure 86 – Prediction Error of Simulated Data

3.2.5 Model Comparison

An adaptive ARMA+RLS model is used on both experimental and simulated data as a comparison. In this section, we compared the multi-iteration training of the RLS algorithm with the proposed VFF-RLS algorithm. For the multi-iteration RLS, each individual data point was used multiple times in the RLS algorithm to update the coefficients of the ARMA model. The results show that the error of prediction can be reduced significantly through multi-iteration data training by 79.1%. The result of the different iterations of the training is shown in Table 11. It can be seen that the increment in the number of iterations significantly reduces the error in terms of mean square error (MSE), mean absolute percentage error (MAPE), and maximum error. The error converges to a satisfactory value defined by the user after fifteen iterations. However, the added computation time may result in hardware challenge and extra memory required in practical applications.

Table 11 – Error Comparison of Different Iteration Number

	MSE	MAPE	Maximum Error	Percentage of Change
Iteration n=1	**0.0164**	**1.26%**	**3.18**	**N/A**
Iteration n=5	0.0063	0.96%	2.64	17.0%
Iteration n=15	0.0020	0.64%	1.35	48.9%
Iteration n=50	0.0001	0.17%	0.43	68.1%
Iteration n=100	5.47×10^{-6}	0.02%	0.09	79.1%

The gradient based VFF-RLS algorithm overcomes the disadvantage of the multi-iteration RLS algorithm by optimizing the RLS convergence speed with the changes of the forgetting factor. The step size μ is selected to be 0.4 by trial and error, and the upper limit

and lower limit of the forgetting factor is set as 0.995 and 0.8. The comparison result can be seen in Table 12. The VFF-RLS, in comparison with RLS, has marginal improvement during single data iteration in a time-varying environment, as shown in Table II. However, with multiple iterations, the VFF-RLS yields smaller error than the traditional RLS algorithm. The error is reduced more than 50% in terms of the MSE and 42% in terms of the maximum error.

Table 12 – Error Comparison of VFF-RLS and RLS

	VFF-RLS	RLS	VFF-RLS (iteration=5)	RLS (iteration=5)
MSE	0.0850	0.1478	5.47×10^{-5}	1.85×10^{-4}
MAPE	1.4775	1.5509	0.1228	0.1927
Max Error	10.0383	10.4087	0.2920	0.5034

The change of forgetting factor can be seen from Figure 87. The result of Figure 87 indicates that a large or constant forgetting factor does not guarantee the maximum speed of convergence and minimization of error. Instead, the forgetting factor varies all the time during the VFF-RLS algorithm. Table 13 demonstrates the improvement by using the VFF-RLS over the RLS. The VFF-RLS improves the convergence speed by ten times and yields a smaller maximum error for the bearing prognostic model.

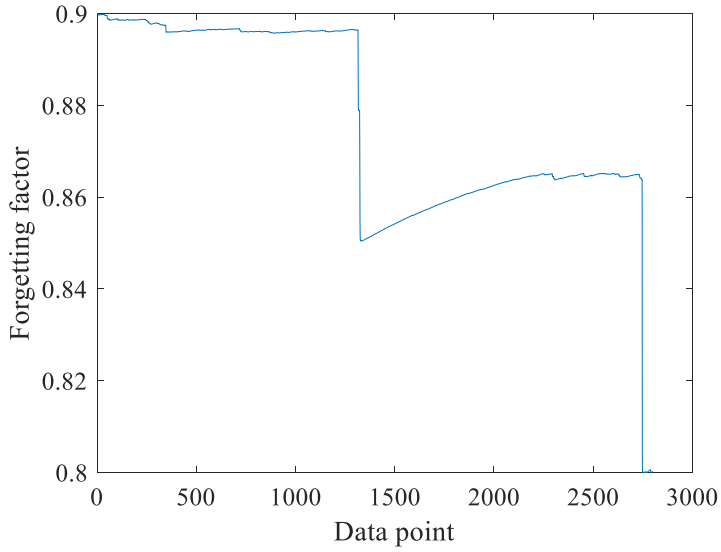


Figure 87 – Forgetting Factor Change During Adaptation

Table 13 – Comparison of RLS and VFF-RLS

	VFF-RLS	RLS
Iteration	10	100
Max Error	0.078	0.09

The ARMA+VFF-RLS model is also compared with a widely used time-series forecasting model, the ARIMA model. The structure of the ARIMA model is selected to be (2,1,2) based on the AIC. The value of the prediction is shown in Figure 88 and 89 along with the prediction from the ARMA+VFF-RLS model.

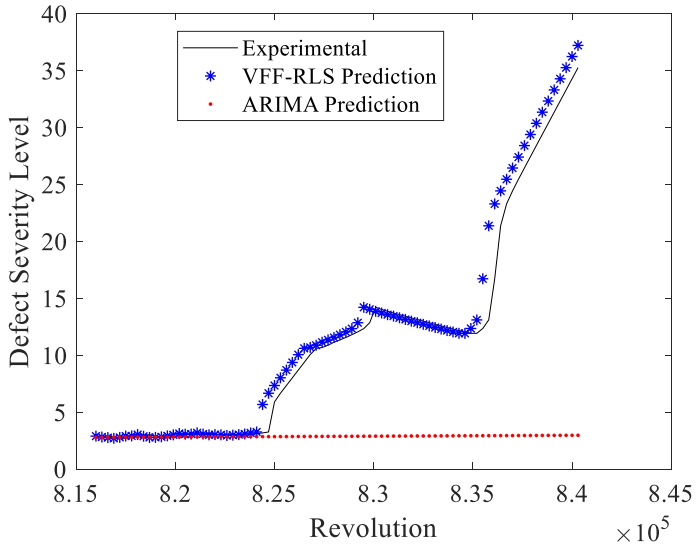


Figure 88 – Comparison of Prediction of Different Models

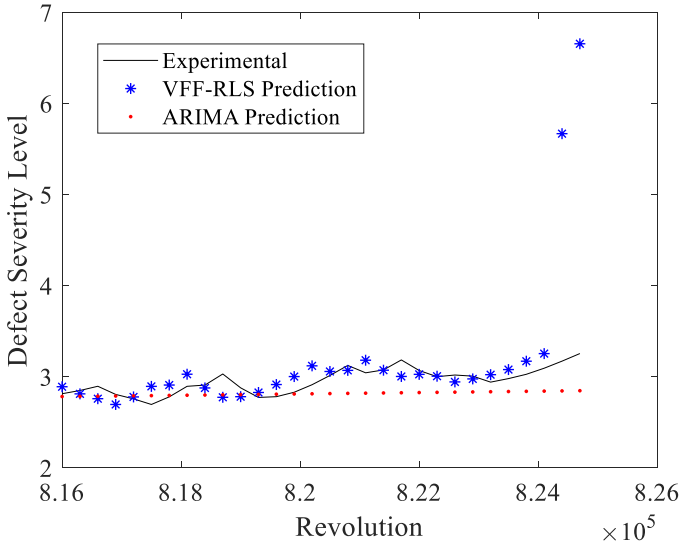


Figure 89 – Comparison of Prediction of Different Models for Stationary Signal

It can be observed that the ARIMA model is unable to adapt to the non-stationary data. In addition, the error of the ARMA+VFF-RLS method is smaller in the stationary part of the signal. In conclusion, the ARMA+VFF-RLS model is superior than the time-series ARIMA model in predicting both stationary and non-stationary trends.

3.2.6 Conclusions

This section presents an adaptive prognostic model that is implemented in online monitoring processes of bearings to predict the degradation trend of bearings. The model first establishes a linear relationship between the vibrational signal and defect-severity level of the bearing. The prognostic model uses an ARMA (2,2) model to predict the defect-severity level of the bearing with the VFF-RLS adaptation. The adaptive model overcomes the inflexibility of the deterministic model by constantly varying the coefficients of the ARMA model. Previous documented models only account for the increase of the magnitude in the vibration signals of bearings. In comparison, this model can accommodate various types of signals and patterns of degradation during the online monitoring process. The VFF-RLS further improves the RLS by increasing the model convergence speed while reducing the error of prediction. The model also yields a better prediction over a widely used ARIMA model which is used as a benchmark model for comparison.

Both experimental and simulated data were tested to verify the applicability and robustness of the model. The adaptive model effectively predicts the defect propagation process by optimizing the coefficients of the ARMA model. The error between the experimental value and predicted value is reduced significantly by the adaptation. The VFF-RLS further optimizes the forgetting factor to minimize the MSE and increase rate of convergence. The iteration number of training data is closely related to the accuracy of the prognosis. A large iteration number generates small error; however, it requires more computation power and memory. The proposed VFF-RLS overcomes the disadvantage of the slow convergence, which resulted from a large forgetting factor.

Since the AIC was used initially in the early stage of bearing vibration signal, the ARMA model structure could change over time during the process of degradation. Further research can be performed to add structural adaptation mechanisms to modify the time-series model. In addition, other signal features such as energy ratio could be combined with the proposed model to predict the trend of degradation.

3.3 PHYSICS-BASED INTELLIGENT PROGNOSIS FOR ROLLING BEARING WITH FAULT FEATURE EXTRACTION

3.3.1 Introduction

In the area of rotating machineries, a very common problem in the operation of gearboxes, electric motors and machine spindles is the unpredictable failure of the bearings [226]. The success of implementing CBM for rotating machineries heavily depends on the prediction of running conditions of the bearings. The new bearings are initially running in smooth conditions without additional vibrations and degradation. However, as the bearings run for a long time, the stability of the system is disturbed by the fatigue behavior of the material made of the rotating elements of the bearings [227]. The deterioration of the mechanical components of the bearings allows cracks within the bearings to grow in an unpredictive manner [228]. In addition, the failure of the bearing does not follow any specific trend [4]. Hence, the randomness of the degradation process significant increases the difficulty of predicting bearing failure and the cost of proactive maintenance of the machinery. Therefore, methods that can capture the bearing failure are still under development for decades to reduce the potential risk of the failure of the bearings.

Previous bearing prognosis related research can be categorized into a few types. The early model includes The Lundberg and Palmgren (LP) model which predicts the fatigue life and load ratings of rolling bearings [201]. However, this model under-estimates the fatigue lives of the bearings and it is hard to implement without prior experimental data. Later, Harris and Yu [229] modified the LP model with considerations of failure stress and stressed volume. However, the hard-to-estimate model parameters prevent the applicability of the model under actual industrial applications. Another category is based on the Paris' Law [7] which, in general, requires measurement of the defect size. Liang *et al.* [73] combined the Paris' law with the nonlinear Recursive Least Square (RLS) algorithm to predict the spall area size. However, the model relies on experimental data of the defect size. In industrial applications, the defect size has to be measured by disassembly and assembly of bearings tested on specialized equipment [7]. The measurement takes significant amount of time and cannot be performed without interruption of the experiment. The last type of prognostic models involves artificial intelligent methods such as Neural Network (NN) and Statistical Model to predict the remaining life of bearings. Gebraeel *et al.* [230] uses a NN based approach to predict residual life of the bearings. However, the time of prediction in this literature is made at the late stage of degradation of the bearings. Kim *et al.* [136] implemented support vector machine to obtain the health states and predicted the remaining useful life. The method requires a large amount of data of bearing failure test, which may not be applicable in certain application. Soualhi *et al.* [231] used Hidden Markov Models and Adaptive Neuro-Fuzzy Inference System to predict bearing failures. However, this method still requires a significant amount of experimental data and it was only applied to the degradation patterns following the Paris' law. Therefore, a

method that predicts the failure of bearings of different failure patterns in the early stage of degradation is needed.

Realized Volatility (RV) has shown great utility in detecting sudden changes in time series for the last few years in various area [232-234]. RV turns out to be a reliable indicator for sudden changes in time-series data displaying long-range dependence [232]. Recent models in financial research has implemented RV with Auto-Regressive Conditionally Heteroscedastic (ARCH) to forecast financial data [235]. Shirota *et al.* uses RV in combination with autoregressive fractionally integrated moving average model to improve the forecasting accuracy of S&P 500 returns data [236]. Therefore, this manuscript implements RV as a tool for the detection of the degradation trend of rolling element bearings.

Li *et al.* investigated different vibrational characteristics of the acceleration data of bearings for fault features extraction. He concluded that general features such as Root Mean Square (RMS), Kurtosis and peak value may not be good estimators for the faulty signal of bearings [237]. Instead, he proposed that the energy ratio should be used to accommodate for both normal and abnormal degradation patterns of bearings. In addition, the general features calculated from the vibrational signal of bearings do not reveal the faulty signal in the early degradation stage. Instead, the energy ratio indicates weak fault signal in the early stage of the degradation of bearings. Therefore, it is a better indicator for weak fault versus the traditional calculated indicators such as the RMS and Kurtosis.

In comparison with the traditional NN, Wavelet Neural Network (WNN) uses nonlinear wavelet basis functions instead of the sigmoid activation function of NN to

improve fitting to nonlinear problems and adapt to high frequency non-stationary signals [238-240]. Several researches have demonstrated the advantages of WNN in prognosis [241-243] in enhancement of prediction accuracy and reduction of unpredictability. However, the prognosis using intelligent algorithm purely relies on statistical analysis without consideration of real time running condition of bearings. The lack of physics of the documented intelligent algorithm has motivated researchers to develop models with physics embedded into the intelligent model. For instance, Lu *et al*. develop an intelligent algorithm based on the physical understanding of the electro-chemical machining process [244]. With the physics included in the intelligent model, more accurate prediction can be achieved.

This section presents a physics-embedded intelligent bearing prognostics model to predict the failure cycle of bearings in the early stage of degradation. The model utilizes the physical properties of the bearing such as the load ratio, rotating speed as part of the input. The prediction point is determined by obtaining the maximum RV value. The energy ratio is used to address the deficiency of existing models in predicting degradation patterns of bearings. The WNN model offers better prediction over the traditional NN model and increases the accuracy in predicting stochastics and random components. The prediction accuracy of the proposed model is compared with a widely used time-series model to demonstrate the advantages of the proposed model.

3.3.2 Prognostic Model

3.3.2.1 Physics-based Model and Parameters

To address inability for early defect detection and lack of physical understanding in the documented prognostic models for bearings, this manuscript developed a prognostic model using WNN to predict the failure cycles of bearings running under various load and speed. The maximum RV value is calculated to detect the fault initiation and summation of the Energy Ratio until the fault initiation point is used as the input to the neural network. Based on the equation estimating bearing fatigue life proposed by Harris [245], the ratio of the actual load versus the bearings' dynamic load is calculated as another input. In addition, the running speed of the bearing and the running cycles where the RV values are the highest are used as input values to the WNN. The WNN was trained using 14 sets of experimental data of run-to-failure signal of the bearings. After training, the WNN makes prediction based on other sets of truncated experimental data. The purposed model is shown in Figure 90.

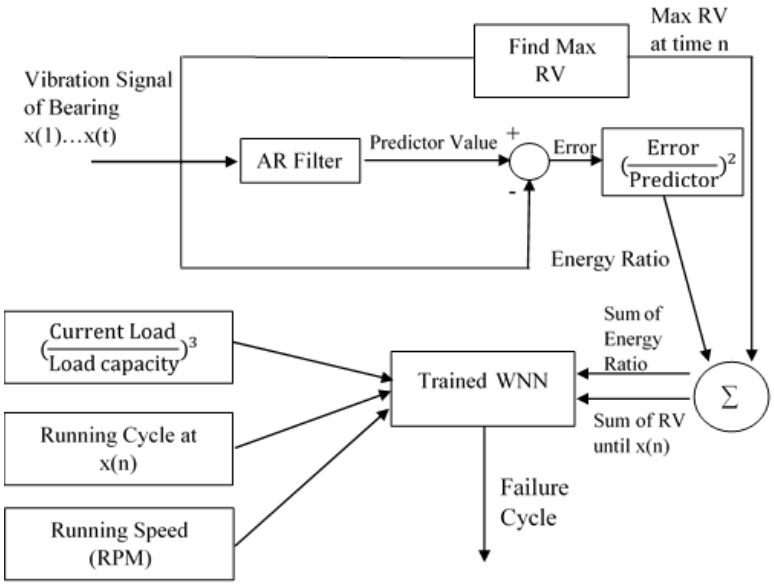


Figure 90 – Prognostic Model of Failure Cycles

Three of the input values of the WNN can be calculated directly from the experimental setup. These values are the load ratio, running cycle at the prediction point and the rotating speed of the bearing as shown on the lower left of Figure 89. To obtain the energy ratio, the vibrational signal of the bearing is passed through an Auto Regressive (AR) filter. In the meantime, the corresponding RV values of all the vibrational signal is calculated until a maximum RV value reaches a threshold value. The sum of the energy ratio and RV are calculated as two inputs for the WNN. Then, the trained WNN makes prediction of the failure cycle based on the inputs. The next section describes the details of obtaining the input values of the WNN.

The first part of the physics-based model is concentrated on signal processing. The vibrational signals of the tested bearings are sent to an AR filter to separate deterministic data and random noisy data shown as error in Figure 89. The AR filter is generally represented by (148):

$$x_t = C_0 + C_1 x_{t-1} + \varepsilon \tag{148}$$

where x_t is called the predictor value, C_0, C_1 are two constants, x_{t-1} is the previous value, known as the lag 1 value, and ε is the prediction error. Sawalhi *et al.* [246] states that the random error part, presented by ε in (148), of the vibrational signal includes the faulty signal of the bearings. This random error is often neglected during the vibrational analysis of bearings. Li *et al.* [237] points out that the energy ratio define in (149) is the critical characteristics to track the degradation of bearings. They believe that as the defect propagates within the bearing, the amplitude of the random noisy part of the bearing data will increase while the deterministic part of the data remains relative stationary. The error in (149) is equivalent to ε in (148), and the predictor in (149) is equivalent to x_t in (148).

$$Energy\ Ratio = \left(\frac{Error}{Predictor} \right)^2 \tag{149}$$

The order of the AR filter in this section is selected to be 1 for the simplification of the computation. The AR filter coefficient is determined to minimize the least square error using the Yule-Walker method [247].

From the fundamental of fracture mechanics, the rate of crack propagation is related to the energy release rate [167]. Once the accumulated energy reached a critical level, the crack growth is unstable. It is believed that the degradation of bearings follows the same principle of fracture mechanics [248]. Camci *et al.* [249] mentioned that the indicators of bearing failure from the horizontal vibrational signal is easier to distinguish than from the

vertical signal. Therefore, this manuscript uses the horizontal vibrational signal to calculate the RV value according to (150):

$$RV_i = 100\,(\log(x_i) - \log(x_{i-1})) \tag{150}$$

where x_i is the vibration data at time equal to *i*. The RV series can be constructed by calculating each RV value. The vibrational data from the experiment performed by Nectoux *et al.* [207] is used to obtain the RV series. The rotational speed of the ball bearing is 1800 rpm and the applied load is 4000 N. A sample of RV value is shown in Figure 91. for a set of experimental data. The corresponding RMS value of the vibrational data of the bearing is plotted in Figure 92. By searching for the maximum RV value and its corresponding cycle number, the sudden change of the vibration can be identified. The point where RV reaches the maximum is set as the point to forecast the failure cycle.

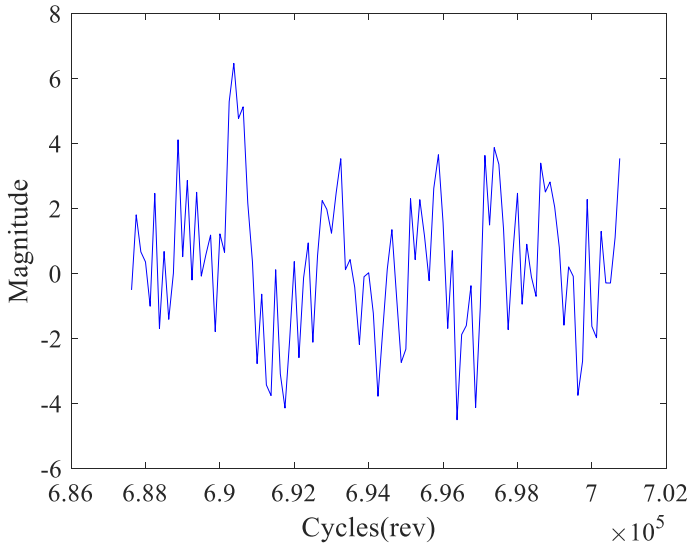


Figure 91 – RV Value of Vibrational Data

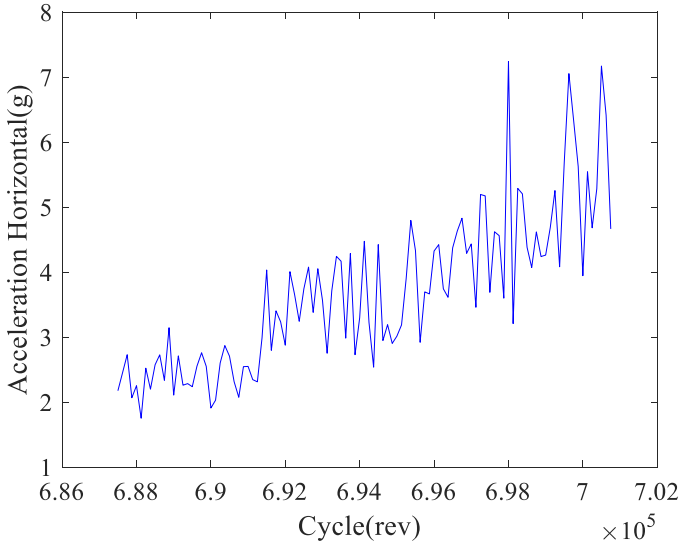


Figure 92 – RMS Value of Vibrational Data

Harris [245] states the fatigue life of a bearing considering point contact can be approximated by (151), where a is a constant, Q is the load applied when the bearing is running continuously and Q_0 is the load capacity of the bearing. Therefore, the cubic value

of the ratio of the current load to the load rating of the bearings is calculated and used as an input to the WNN. The rotational speed and the cycles at which the RV value reaches a maximum, as shown in Figure 90, are used as inputs to the WNN model,

$$L = a\left(\frac{Q_0}{Q}\right)^3 \tag{151}$$

3.3.2.2 WNN Configuration

In the recent few years, wavelet neural network has shown great applicability in analysis of data and forecast of time series. Doucoure *et al.* [250] used wavelet neural network in predicting wind speed data. They stated that the WNN algorithm developed eliminates the component with low predictability in order to optimize the performance. Okkan [251] implemented a WNN model to predict the monthly reservoir inflow. The Discrete Wavelet Transform (DWT) in the algorithm erases the ineffective sub-time series to enhance the accuracy of the prediction. The wavelet coefficients are generally denoted by (152) where a is the scaling factor, τ is the time shift, and $\psi(t)$ is generally called the mother wavelet constructed from translations by τ and dilations by a. The scaling factor can be adjusted to obtain different frequency components for varies applications.

$$C(a,\tau) = \frac{1}{\sqrt{a}}\int f(t)\psi\left(\frac{t-\tau}{a}\right)dt \tag{152}$$

The inverse transform of the wavelet, which reconstructs the original signal, is shown in (153). C_ψ is a wavelet function dependent constant and $\psi(t)$ is the wavelet function.

$$f(t) = \frac{1}{C_\varphi} \int_a \int_\tau < f(t), \psi_{a,\tau}(t) > \psi_{a,\tau}(t) d\tau \frac{da}{a^2} \tag{153}$$

A wide variety of wavelet function, $\psi(t)$, exists for different applications. In this section, the Morlet Wavelet is used in the WNN because its effectiveness in signal processing and network estimation. The Morlet Wavelet applied to the WNN is shown in (154),

$$\psi(t) = \exp(-t^2 / 2) \cos(1.75t) \tag{154}$$

The WNN used in this section is shown in Figure 93. It consists of three layers, one input layer, one hidden layer and one output layer. The number of hidden neurons is selected to be 6 by trial and error. The learning rate from the input layer to the hidden layer and the hidden layer to the output layer is select to be 0.1 and 0.01, respectively. The number of epochs is selected to be 100 to reduce the possibility of overfitting. The summation of the wavelet function equals to the weight of the WNN as shown in Figure 94.

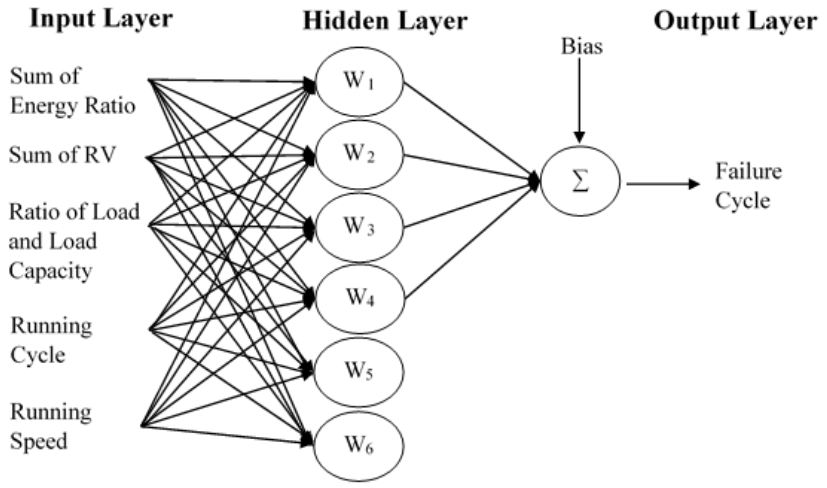


Figure 93 – WNN Structure to Predict Failure Cycle

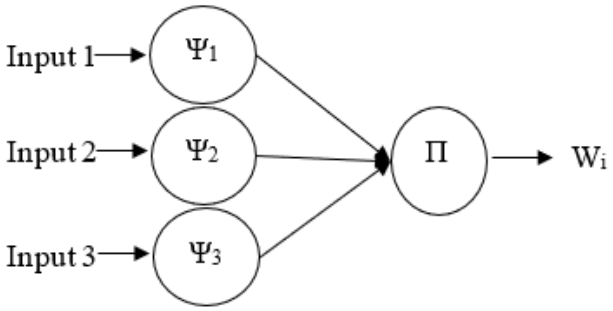


Figure 94 – Summation of Wavelet as the Weights of the WNN

The output of the *i*-th layer of the wavelet neuron is denoted by (155), where x_i is the i-th input and *j* represents the number of wavelet neuron. τ and *a* are the translation and dilation, respectively.

$$\psi_j(x) = \sum_{i=1}^{n} \exp\left(-\left(\frac{x_i - \tau_j}{a_j}\right)^2 / 2\right) \cos\left(1.75 \frac{x_i - \tau_j}{a_j}\right) \tag{155}$$

The output of the output layer shown in (156) is the weighted sum the wavelet function $\psi(x)$.

$$y(x) = \sum_{i=1}^{n} w_i \psi_i(x) \tag{156}$$

The training of the WNN allows its weights to be updated, which provides accurate result. The weight adjustment relies on the minimization of the cost function defined by (157) and (158):

$$Error = \frac{\sum_{i=1}^{n} e(i)^2}{2n} \tag{157}$$

$$e(i) = x(i) - \hat{x}(i) \tag{158}$$

where $e(i)$ is the difference between actual value $x(i)$ and predicted value $\hat{x}(i)$. The parameters w_i, τ_j, a_j are updated according to (159) to (161):

$$\Delta w_i = \frac{\partial Error}{\partial w_i} = \frac{\partial Error}{\partial y} \frac{\partial y}{\partial w_i} \tag{159}$$

$$\Delta \tau_j = \frac{\partial Error}{\partial \tau_j} = \frac{\partial Error}{\partial y} \frac{\partial y}{\partial \psi_j} \frac{\partial \psi_j}{\partial w_i} \frac{\partial w_i}{\partial \tau_j} \tag{160}$$

$$\Delta a_j = \frac{\partial Error}{\partial a_j} = \frac{\partial Error}{\partial y} \frac{\partial y}{\partial \psi_j} \frac{\partial \psi_j}{\partial w_i} \frac{\partial w_i}{\partial a_j} \tag{161}$$

The weight of the WNN updates according to (162), where $w(i+1)$ is the newly updated weight, $w(i)$ is the previous weight, η is the learning rate and α is a parameter related to the convergence.

$$w(i+1) = w(i) + \eta \Delta w(i) + \alpha \Delta w(i-1) \tag{162}$$

3.3.3 Experiment

The purpose of the experiment is to demonstrate various trends of degradation of the bearings under either the same or different loading conditions and variable rotational speed. The experimental data of the bearings are obtained from Nectoux *et al.* [207]. The dynamic load capacity of the bearing is rated at 4000 N. Three test conditions were used at 1800 RPM with 4000 N load, 1650 RPM with 4200 N load and 1500 RPM with 5000 N load respectively, and the three conditions are designated as 1, 2, 3 of the first digits of the bearing number in Table 14. The second designation indicates the number of the exact same bearings used in the experiment. The failure of the bearings is defined as the time when the acceleration recorded exceeds 20 g by Nectoux *et al.* [207]. The failure cycles of the bearings are also provided in Table 14 to facilitate the model training and prediction. The failure cycle is calculated by finding the cycles where the acceleration first exceeds 20 g.

Table 14 – Running Load and Failure Cycles of the Bearings

Bearing Number	Load (N)	RPM	Failure Cycle
1-1	4000	1800	829688
1-2	4000	1800	261300

Bearing Number	Load (N)	RPM	Failure Cycle
2-1	4200	1650	248910
2-2	4200	1650	217430
3-1	5000	1500	123410
3-2	5000	1500	409250
1-3	4000	1800	712500
1-4	4000	1800	351870
1-5	4000	1800	738900
1-6	4000	1800	734400
1-7	4000	1800	677700
2-3	4200	1650	537625
2-4	4200	1650	206525
2-5	4200	1650	635525
2-6	4200	1650	192775

Bearing Number	Load (N)	RPM	Failure Cycle
2-7	4200	1650	63250
3-3	5000	1500	108500

3.3.4 Results

Figure 95 shows the original vibrational signal of a bearing under abnormal degradation pattern. The general features such as RMS, Kurtosis and Skewness value were compared with the Energy Ratio to demonstrate the superiority of using the Energy Ratio as a primary indicator for degradation. Figure 96 to 98 show the raw signal, RMS, Kurtosis and Skewness of the bearing tested in the experiment. It is difficult to predict any trend from the figures because the signal exhibits abnormal patterns of degradation. The RMS value initially decreases and because of the decreasing value, a time-series forecasting model such as Auto Regressive Integrated Moving Average (ARIMA) and Crack propagation model will not work properly in forecasting by using the RMS. In addition, the kurtosis value does not generate valuable information as well as the Skewness value. However, by calculating the energy ratio using the AR filter, a normal bearing degradation pattern can be obtained as shown in Figure 99. As the running cycle number of the bearing increases, the degradation trend reflected by increasing energy ratio can be observed from Figure 99. The Energy Ratio data can be treated as time series and various models can be implemented to predict the degradation trend of the bearings.

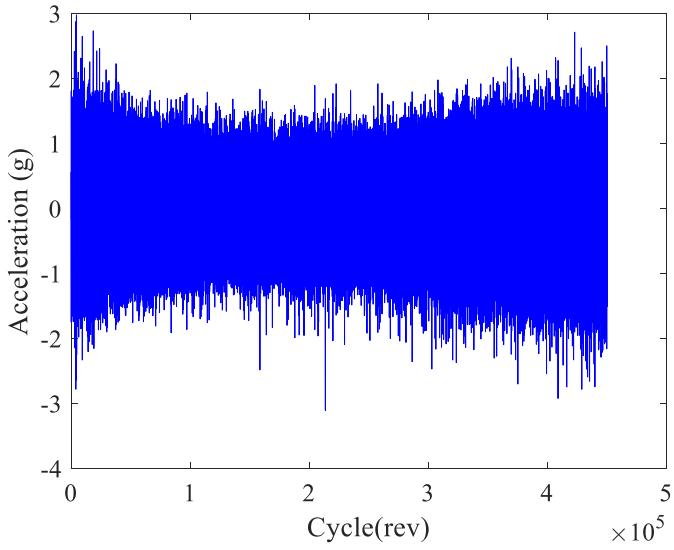


Figure 95 – Unprocessed Signal of Abnormal Degradation Pattern of Bearing

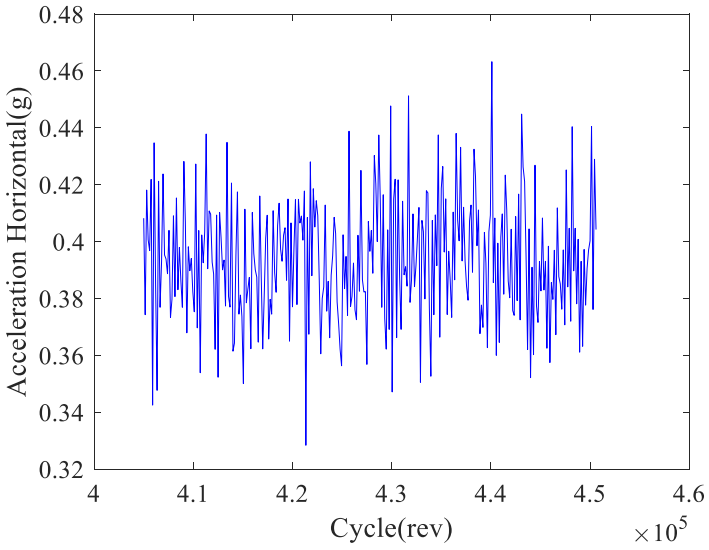


Figure 96 – RMS Value of the Bearing

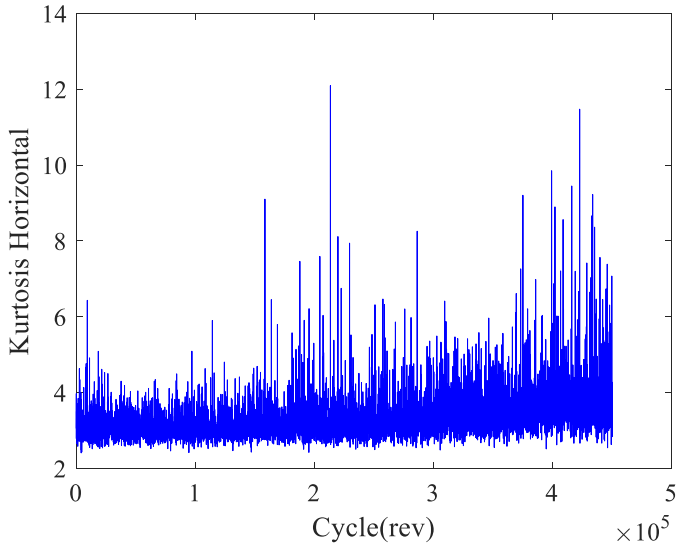


Figure 97 – Kurtosis Value of the Bearing

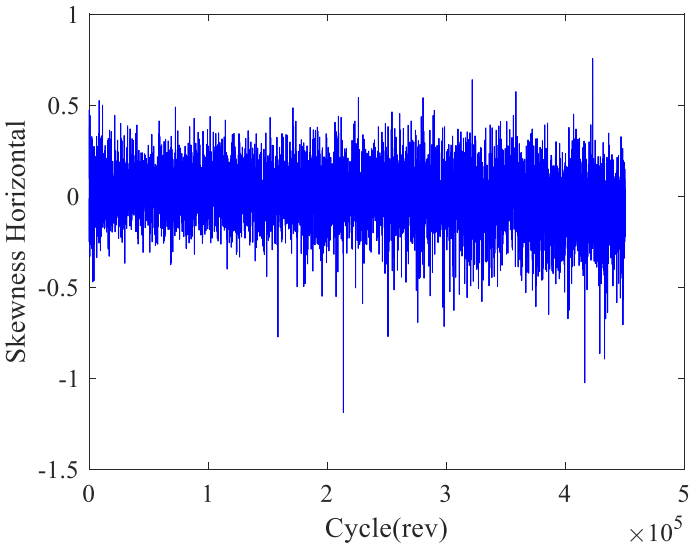


Figure 98 – Skewness Value of the Bearing

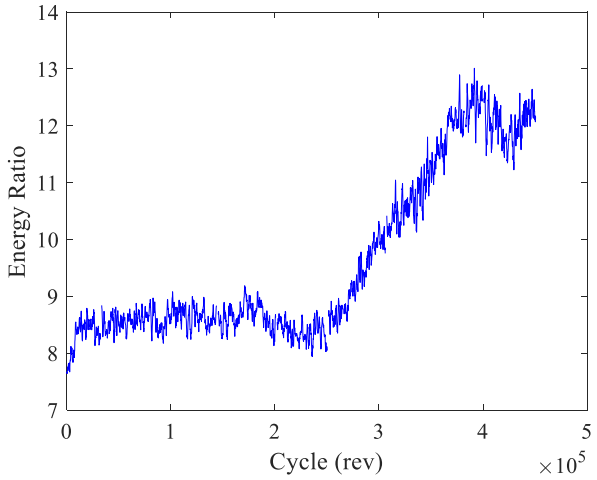


Figure 99 – Energy Ratio of the Bearing

To demonstrate the applicability of forecasting degradation trend using the Energy Ratio in both normal and abnormal patterns of the signal, another set of bearing data with normal degradation pattern is tested, and the Energy Ratio is plotted in Figure 100. The degradation pattern of the bearing which always has an increasing trend in the Energy Ratio can be easily approximated by various models such as an Auto-Regressive Integrated Moving-Average (ARIMA) model. By examine Figure 99 and 100, it can be seen that because of the overall increasing trend of the Energy Ratio, it is a good indicator of the degradation patterns of bearings in both normal and abnormal patterns.

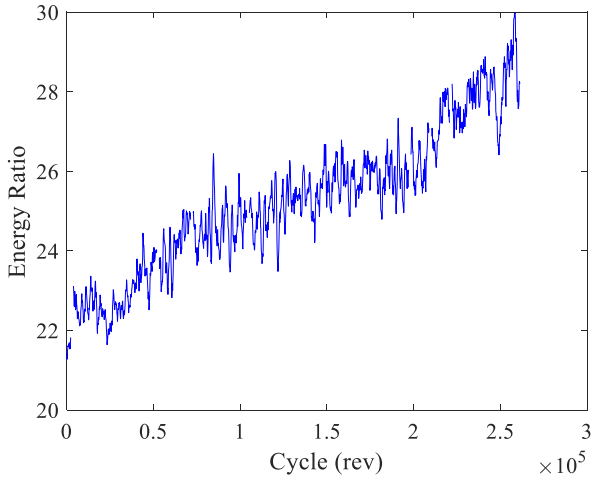


Figure 100 – Energy Ratio of the Bearing

Rare cases which invalidate using the Energy Ratio alone as an indication of the failure of bearings do exist as shown from Figure 101 to 102. Figure 101 and 102 demonstrate a situation where the Energy Ratio alone may not be a good indicator of failure. It can be seen from Figure 101 that the vibrational signal shows an increasing trend in the beginning but decreases to a stable trend later. The corresponding Energy Ratio shown in Figure 102 exhibit the same trend. Because of this abnormal decreasing trend in vibrational signal while the defect is growing, the energy ratio alone may not be used directly to forecast degradation trend of bearings.

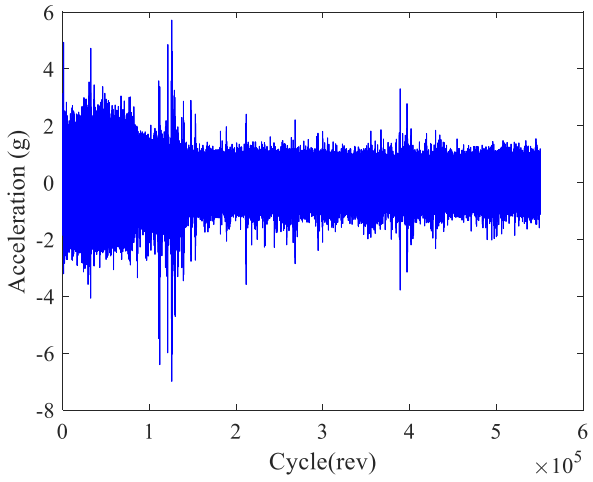


Figure 101 – Abnormal Degradation Pattern of Bearing

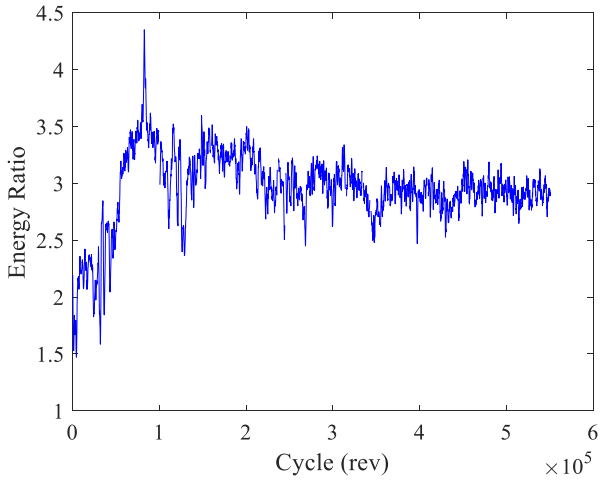


Figure 102 – Energy Ratio of the Abnormal Degradation Pattern

To resolve the issue of the non-increasing trend in the energy ratio, this manuscript proposes using the sum of the energy ratio as an input to the WNN. The reason behind using the summation of energy ratio is based on the understanding of fracture mechanics. Skelton *et. al* states that the process zone near a crack tip fails when the accumulated energy within the zone reaches a critical value [252]. Therefore, the sum of the energy ratio is considered closely related to the failure of the bearing. Before deciding how many data

needed to be summed up, the RV value of the Energy Ratio is calculated. The RV value corresponding to Figure 100 is shown in Figure 103. The maximum RV value which indicates a sudden change of the variation of the signal is at 199240 cycles in Figure 103. Therefore, we calculate the cumulative sum of the energy ratio and the RV value until the 199240 cycles and use those values as the inputs for the WNN.

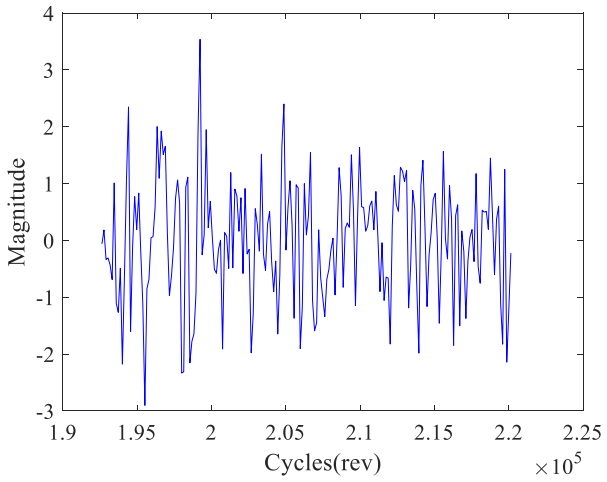


Figure 103 – RV Values of Energy Ratio

The truncated bearing data 1-3, 2-2, and 3-1 which represents three different experimental running conditions were used for prediction and the rest of the data were used for training the WNN. The prediction result can be seen in Table 15. In addition, another set of the training and testing data is tested for the WNN as shown in Table 16. The data with the same first designation number of the bearings used in Table 15 and 16 are two sets of data using the same bearings and experimental setup. By comparing the failure cycle of bearing 2-2 and 2-3, we conclude that although the bearings and experimental setup are the same, the failure cycle of the bearings could be significantly different. This type of intrinsic

property of bearings increases the difficulty of forecasting the failure cycles and predicting remaining useful life.

By comparing the data in Table 15 and 16, the predictions using the proposed WNN model yield satisfactory result for most of the data based on the limited amount of training data. The error could be reduced by increasing the training iteration. However, care must be taken to prevent the overfitting of data. The largest over-prediction as shown in Table 15 and 16 for bearing 1-3 results in a predicted failure time 10.5 min after the bearing reaches the 20 g failure criteria. For bearing 2-3, the model predicts an early failure of 13 min before the vibrational signal reaches the 20 g level. The large percentage error is generated in the bearings of the second running condition. The cause are attributed to the large amount of data truncated from the experimental data available in [207]. With sufficient vibrational data and optimized WNN structure, the prediction error can be further reduced.

Table 15 – Prediction Result

Bearing	Actual Failure Cycle	Predicted Failure Cycle	Error Percentage
1-3	712500	902942	26.7%
2-2	217430	317961	46.2%
3-1	123410	145774	18.1%

Table 16 – Prediction Result

Bearing	Actual Failure Cycle	Predicted Failure Cycle	Error Percentage
1-6	734400	699125	4.8%
2-3	537625	323780	39.8%
3-3	108500	122854	13.2%

The prediction result of the proposed WNN model is compared with an ARIMA model. The ARIMA model coefficients are estimated using the Yule-Walker method as in [205] and the model parameter is determined by implementing the Akaiki Information Criterion. The ARIMA model was first trained using the nearest 500 data points of the RMS value of the horizontal acceleration at the point where the RV value exceeds the threshold. The iterative prediction is implemented to predict future value of the vibration of the bearings. The cycle where the vibrational signal of the bearing reaches the failure threshold is located and compared with the result of the WNN model. The result of the comparison of the prediction of the ARIMA model and the WNN model is shown in Table 17.

Table 17 – Model Comparison

Bearing	WNN Predicted Failure Cycle	Error Percentage	ARIMA Predicted Failure Cycle	Error Percentage
1-3	902942	26.7%	465300	34.7%
2-2	317961	46.2%	217663	5.2%
3-1	145774	18.1%	6375	94.8%
1-6	699125	4.8%	772500	5.2%
2-3	323780	39.8%	1532025	185.0%
3-3	122854	13.2%	156375	44.1%

The WNN model outperforms the widely used ARIMA model in the accuracy of prediction. In the case of bearing 3-1 and 2-3, the ARIMA model was not able to follow the bearing degradation trend due to the irregular bearing degradation pattern as shown in Figure 95 and 101. However, the RV-WNN model overcame the difficulty of various degradation patterns and was able to make better prediction before significant damage of the bearings is detected.

3.3.5 Conclusions

In this section, a methodology using a WNN model combined with physical understandings of the bearing failure to predict failure cycle is developed. An AR filter was used to obtain the residual of the signal, and the ratio of the residual to the AR filter predictor is defined as the Energy Ratio. The Energy Ratio demonstrates wider applicability in forecasting over the RMS, kurtosis and skewness. The accumulated Energy Ratio is calculated based on the fundamental of fracture mechanics. The RV is used as an indicator for detecting drastic changed in the vibrational signal before significant damage of the bearing is observed. The maximum RV defines the point where the cumulative energy ratio and RV are calculated. The ratio of applied load is related to the failure of the bearing based on previously conducted research. Using the Energy Ratio, RV, ratio of applied load and other parameters as inputs, the proposed WNN predicts the failure cycle within desirable accuracy.

The proposed model overcame the difficulty of documented forecasting models which uses the RMS, Kurtosis and Skewness value of the vibrational signal to make prediction. The WNN model developed also has the advantage of predicting failure cycle in the early stage of degradation process. The Energy Ratio is a better indicator of bearing degradation trend versus the commonly used RMS, Kurtosis and Skewness value in various degradation patterns. The newly developed prognostic model yields better result of prediction of failure cycle than that of the ARIMA model. The WNN model overcomes the incapability of predicting abnormal degradation pattern using the ARIMA model and does not require significant damage detected to make accurate prediction. Future research could incorporate

a failure time envelope to add more statistical meaning to the predicted bearing life and further explain the stochastic nature of bearing life.

CHAPTER 4. OPTIMIZATION OF PROCESS PARAMETER IN ELECTRO-CHEMICAL MACHINING

4.1 INTELLIGENT MODEL FOR PROCESSING OPTIMIZATION IN ECM

Product quality is one of the most important factors in today's competitive manufacturing industry. Achieving the desired quality of the manufactured good while maintaining a relatively low cost has become a common goal for various manufacturers around the world. In the areas of non-traditional manufacturing processes, such as ECM, EDM, and LBM, achieving the desired quality of the final product poses a major challenge as the relationship between the process input parameters and KPIs is not fully understood. Though the mechanism of these processes is not fully grasped, these non-traditional manufacturing techniques are advantageous compared to traditional manufacturing processes. Nowadays, manufacturing industries use a trial-and-error–based approach to select the optimal input process parameters to achieve the desirable product specifications. However, the trial-and-error–based approach is very time-consuming, inefficient and can tremendously increase the manufacturing costs. To overcome these challenges, many researchers have used intelligent techniques to model the relationship between the input process parameters and the KPIs, as intelligent techniques are capable of analyzing, self-learning, apprehending complexities, and they are able to store and analyze large amounts of data to obtain an increased quality of the product while shortening the time-to-market.

The proposed method embeds the partial knowledge in the intelligent prediction models to further increase their prediction accuracy. An NN prediction model embedded with the partial knowledge about the relationship between the input process parameters and intermediate outputs is created. The partial knowledge is utilized by specifying part of the

NN structure as compared to a pure NN structure where the relationship is completely unknown. The proposed methodology is tested on a case study of µ-ECM drilling.

A traditional ECM setup is shown in Figure 104 [244]. The electrolyte velocity was 10 m/s at the outlet of the pump, the electrolyte temperature was 27 °C, the initial gap between the tool and the workpiece was 100 µm, the tool moving distance was 800 µm, the workpiece material was SUS 304, the electrolyte used was 10% wt. NaNO3, the nominal diameter of the hole was 900 µm, and the depth of the hole was 500 µm.

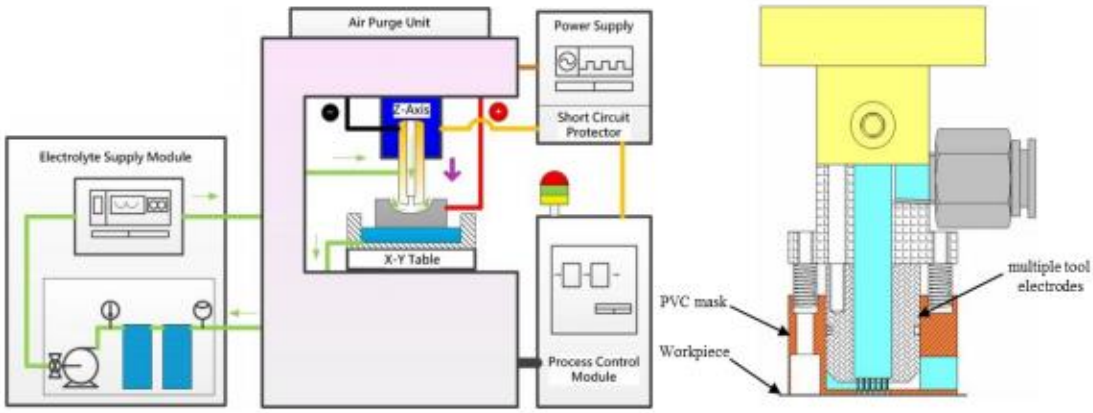


Figure 104 – ECM Setup

The voltage, pulse-on-time, and feed rate were used as the controllable process parameters, while the inner diameter of the micro-hole D_{in} and the outer diameter D_{out} were the measurable performances. To embed the physical knowledge in the NN, the relationship between the inputs and the intermediate outputs was first linearized. This linearized relationship was represented in the NN as linear weighted connections between the inputs and the first hidden layer, where the outputs from the first hidden layer represent intermediate outputs. Additional hidden layers were then added to the NN, which connected the first hidden layer and the output layer where the connections were nonlinear. Once the NN structure with the embedded partial physical knowledge was established, the

weights of the NN could be trained using any gradient-based or metaheuristic algorithm. The prediction model is shown in Figure 105.

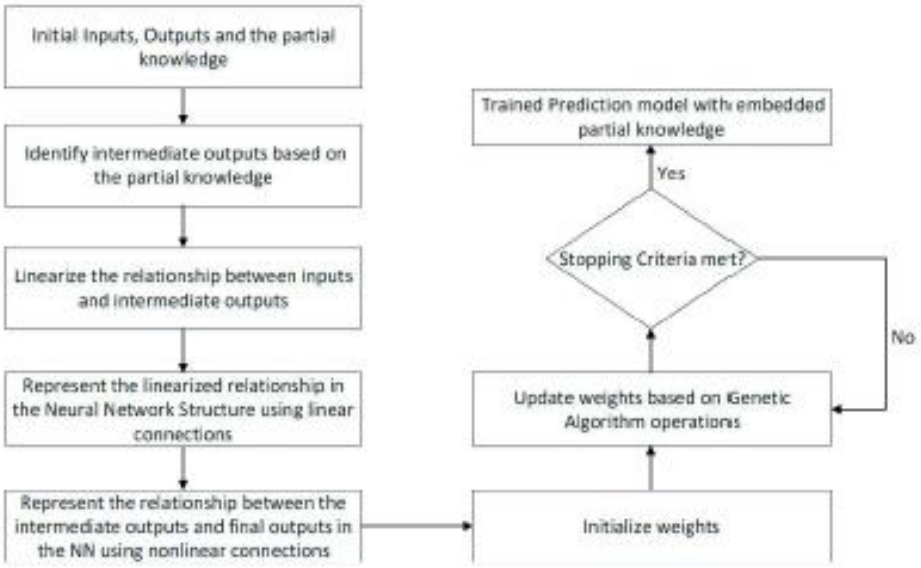


Figure 105 – Physics Embedded Prediction Model

The following assumptions were made to simplify the relationship between the inputs and the intermediate outputs: the electrochemical gap between the tool (cathode) and the workpiece (anode) was assumed to be constant for one set of input parameters (voltage, feed and pulse-on-time). The transient response was ignored, and the whole process was assumed to be in steady state. The electrolyte conductivity was assumed to be constant because the refill tank was significantly large in comparison with the electrolyte involved in the machining process. Therefore, the active ion in the electrolyte could be assumed to be constant. Since the tool was not coated with insulation material, we assumed the electrical field was uniform around the tool. Eddy flow was neglected in the inter-electrolyte gap and the horizontal velocity was assumed to be constant. In addition, friction

and heat transfer were not considered in establishing the physical model due to the high electrolyte flow rate and the low surface roughness.

Four intermediate outputs were determined to be the most influential on the outputs, the current density, the inter-electrode gap, the void fraction, and the material removal rate. According to McGeough [253], the current density in μ-ECM can be calculated using (163):

$$J = \frac{I}{A} = \frac{K_e V}{h} \tag{163}$$

where I denotes the current, A is the area where the current is applied to, K_e represents the electrolyte conductivity which is assumed to be constant, V is the voltage applied, and h is the inter-electro gap width which is set to constant in the assumption. Therefore, the intermediate output J has a linear relationship with the input parameter:

$$J = C_1 V \tag{164}$$

The equilibrium gap size can be described using an exponential function with respect to the pulse time as shown in (165) [254]:

$$h_e = \frac{MVe^{cp}}{f} \tag{165}$$

where M is a machining parameter related to the atomic numbers and valencies of the elements constituting the workpiece, the electrolyte conductivity, Faraday's constant and the density of the workpiece; V is the voltage applied; f is the feed rate of the electrode; c

is a constant; and p is the pulse-on-time. A first-degree Taylor approximation is implemented to simply (166):

$$h_e = C_2 + C_3 V + C_4 f + C_5 p \tag{166}$$

Using the same methodology, the void fraction α can be linearized as:

$$\alpha = C_6 + C_7 V + C_8 f \tag{167}$$

and the material removal rate can be described using:

$$MRR = C_9 + C_{10} f + C_{11} p \tag{168}$$

The structure of the NN is shown in Figure 106.

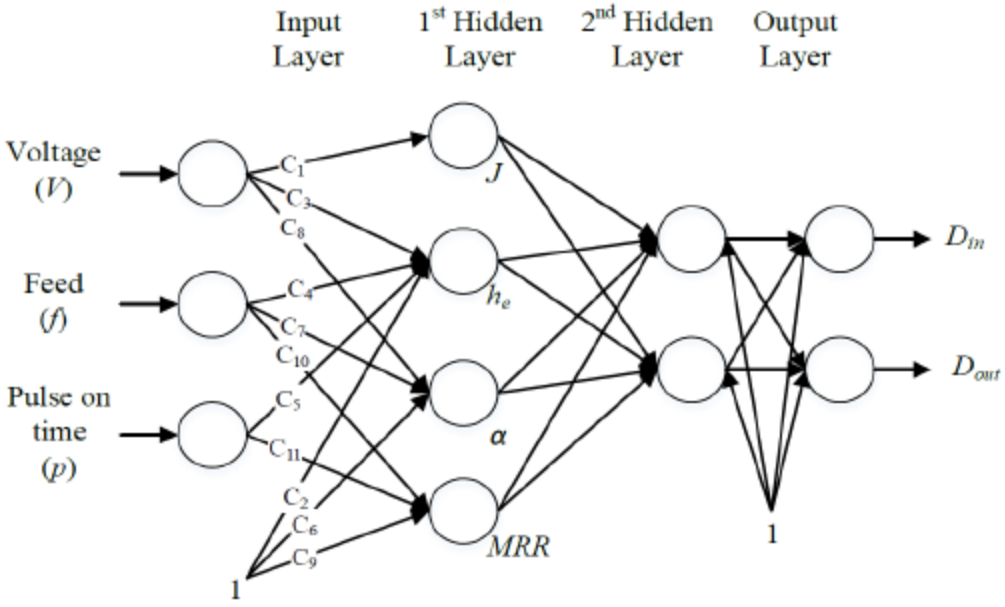


Figure 106 – Physics Embedded NN Structure

The prediction result is shown in Table 18.

Table 18 – Physics Embedded NN Prediction Model Performance

Simulation #	MAPE Obtained Using NN Model with Embedded Knowledge (%)	MAPE Obtained Using Pure NN Model with 3-2-2 Structure (%)	% Improvement
1	8.32	8.87	**6.17**
2	10.01	11.12	**9.99**

Simulation #	MAPE Obtained Using NN Model with Embedded Knowledge (%)	MAPE Obtained Using Pure NN Model with 3-2-2 Structure (%)	% Improvement
3	10.54	11.84	**10.97**
4	8.83	9.07	**2.71**
5	8.69	9.76	**10.92**

Compared to the pure NN model with a 3-2-2 structure, the NN model with embedded knowledge had a better prediction accuracy ranging from 2.71% to 10.97%. The future activity will involve developing improved optimization algorithm to further improve prediction of performance indicators in 3D free form ECM.

4.2 Bayesian Optimized Deep Convolutional Network for Electrochemical Drilling Process

4.2.1 Introduction

Starting from the early casting process to today's hybrid subtractive and additive processing, the increasing demand for product quality has driven the new development of manufacturing processes. Maximizing productivity while reducing cost and achieving a better surface finish that is free of induced residual stress has always been the goal of manufacturing industries. In recent years, the popularity of non-traditional manufacturing processes has increased significantly because of the ability to process difficult-to-cut

material while minimizing mechanically induced stresses. The electrical discharge machining (EDM) process is capable of machining materials with high hardness such as tungsten carbide and carbon fiber-reinforced polymers [255, 256]. However, the recast layer by the electrode deposition and thermally influenced machining zone could affect the material properties [257, 258]. The heat-affected zone (HAZ) will potentially induce residual stresses and phase transformation for heat-sensitive materials [185]. For materials used in critical applications such as nickel titanium alloys the thermally induced phase transformation could significantly influence the proper function of the device and reduce the fatigue life [259, 260]. The laser beam machining (LBM) process is a non-contact process that melts and vaporizes material from the parent material [261]. This process is generally used for the high-precision manufacture of complex shapes. The high-energy beam provides a high material removal rate and is not sensitive to material hardness [262]. However, based on the nature of the material removal process of LBM, it poses the same or an even worse problem as the EDM process. The HAZ could be as large as 3 mm in depth during the LBM process of Ti6Al4V, as reported by Yang et al. [263]. In comparison with the previously mentioned method, the electrochemical machining (ECM) process is capable of processing hard-to-cut material while generating a residual stress-free surface with low roughness [264]. However, the complexity of the ECM process and limited physical understandings have prevented the processing from competing with other processes in mass production.

Lohrengel et al. investigated the anodic dissolution of the ECM process and reported that the oxide film and supersaturated viscous film consist of the dissolved material and depleted ions could affect the material removal [265]. However, these

quantities are not easily measurable during the machining process, and hence are difficult to control. Bhattacharyya et al. examined the influence of current efficiency, power supply, tool design, electro-gap, and electrolyte of ECM [266]. Although it provides insights into the qualitive selection of process parameters, quantitative measures such as the applied voltage, current, and feed rate are important for manufacturers in the setup of the machining process. McGeough presented analytical solutions for several process parameters in [253]. The hydrogen bubble at the inter electro-gap is a key factor for the determination of the material removal rate. Thorpe et al. [267] developed an analytical representation of the void fraction in the electrolyte between the electrode. However, the equilibrium cannot be maintained during the process because of the small inter electro-gap size and insufficient electrolyte supplies. Because of the abovementioned difficulties, industry and research tend to seek help from intelligent techniques which are capable of analyzing highly nonlinear problems through learning from experience. The emerging deep learning algorithm has quickly gained favor across different industries for solving prediction problems.

Zain et al. implemented neural network to predict surface roughness in the machining process using cutting speed, feed rate, coating, and radial rake angle [268]. Various network structures are compared to determine the best selection of network structure using trial and error. Lu et al. [264] developed a neural network with partially imbedded physical understanding to predict drilling diameter for the ECM process. The structure of the network is pre-defined using known analytical solutions; however, the rest of the neural network structure is determined by trial and error. Fu et al. proposed a deep belief network (DBN) to classify the cutting state for machine monitoring in [269]. The network has shown significant improvement in error reduction in comparison to the

traditional neural network and support vector machine. However, the details of selection of the network structure were not discussed. Li et al. implemented DBN to predict machine backlash error using inputs such as machining torque, ambient temperature, and measured position [270]. The trained network yields an accurate prediction for the backlash error. However, the selection of the network structure appears to be arbitrary. Other similar works using learning algorithms are presented in [271]. The lack of a systematic method to select the optimal network structural remains a challenge today. To shorten the time-to-market, a more systematic and scientific approach needs to be developed for neural network hyperparameter optimization.

In this section, a Bayesian optimization algorithm is implemented to optimize structures of a deep convolutional neural network which uses voltage, feed rate, and pulse-on time to predict the ECM drilled diameter at the entry and exit of the hole. The deep learning convolutional neural network consists of one convolution layer, one random drop out layer, and three fully connected layers to map out the relationship between inputs (feed rate, voltage, pulse-on time) and outputs (entry and exit hole diameters). Because ECM is a very complicated process, the implementation of the highly nonlinear deep convolutional network helps improve the prediction accuracy. The Bayesian optimization algorithm aims to find an optimal set of parameters that will minimize the mean square error (MSE) without performing an exhaustive search that demands significantly more computation power. The proposed network is compared with a previously developed physics embedded neural network and a traditional neural network to demonstrate its improved performance. The rest of the section is organized as follows: Section 2 briefly describes the deep convolutional network and Bayesian optimization; Section 3 describes the experimental

setup of the μ-ECM; Section 4 presents the results based on the proposed approach and a comparison with previous work; and Section 5 presents conclusions from the presented work as well as possible future directions.

4.2.2 Deep Convolutional Network Prediction Model for ECM

4.2.2.1 Deep Convolutional Network

The deep convolutional network, as a novel and powerful tool to capture the complex dependency between input and output signals, to the best of our knowledge, has not been implemented in the ECM process. In addition, the networks are generally tuned by trial and error. The Bayesian optimization algorithm facilitates the automatic tuning of the network without human intervention, which significantly reduces the effort or skill level required for industrial applications. In previous documented research, the convolutional neural network (CNN) was demonstrated to work well in identifying patterns in data. Moreover, it has been adopted in various applications in scenarios with the absence of human experts or for the adaption of solutions to specific cases [189-191] because of its capability of solving highly nonlinear problems and its generalization to different data types. It was initially proposed for pattern recognition with back propagation in 2D or 3D applications such as images and video frames [272]. Alternatively, 1D CNN was developed for applications over 1D signals such as electrocardiogram (ECG) and mechanical data. Our 1D CNN architecture has a combination of two types of layers: convolution and fully connected. In the convolution layers, the data are convolved using local learnable kernels to form the output feature maps. For the neural nets with a specified nonlinear function and choice of activation function, the following equation is given:

$$x_{out}^{(l)} = g(h^{(l-1)}) = \sigma(\sum x_{in} \times W_j + b_j) \tag{169}$$

where $x_{out}^{(l)}$ is the output feature map of the current layer l, x_{in} is the input data, and W_j and b_j are the kernel and bias for the current layer.

The output feature maps are then flattened and fed into the next layers for processing. The added convolution is sensitive to nonlinear functions such as sinusoidal and periodic functions—those used in Fourier transform. For the ECM application, because of the unknown intermediate parameters such as the gas bubble fraction, current density, and efficiency, adding the convolution layer could be beneficial to capture their interaction. The softmax function is discarded because the CNN is used for prediction instead of classification. The activation function $\sigma(\bullet)$ between each layer is selected to be a Rectified Linear Unit (ReLU) for computation efficiency [191]:

$$[\sigma(z)]_j = \max\{z_j, 0\} \tag{170}$$

The voltage, feed, and pulse-on time are selected as the input of the network and the drilled hole diameter at the top and bottom are selected to be the output.

The network structure used in the ECM application is shown in Figure 107.

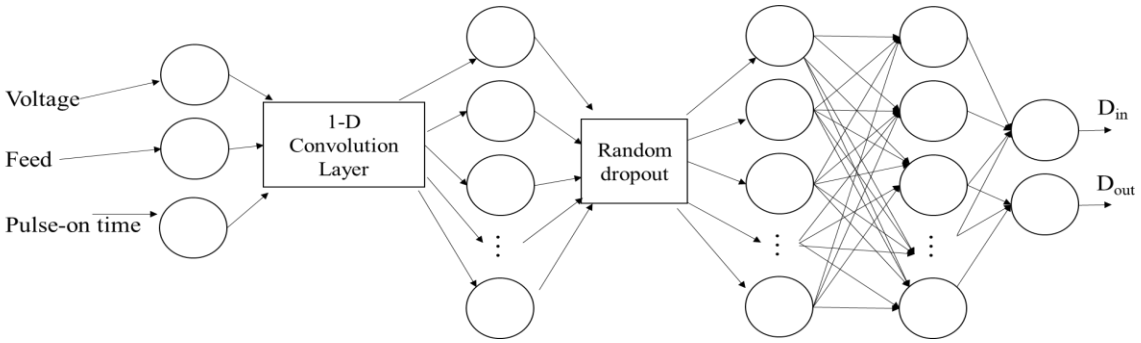


Figure 107 – Deep Convolutional Network To Predict Electrochemical Machining (ECM) Drilling.

A random dropout layer with an optimized percentage of dropout ratio is added between the first and second hidden layer to prevent overfitting [273]. By dropping out subsets of features in the training process, dropout can effectively prevent overfitting. The random dropout for the *l*-th layer can be described as:

$$h_{drop}^{(l)} = h^{(l)} \odot mask^{l} \tag{171}$$

where $\odot$ is the element-wise multiplication and $mask^{l}$ is a vector of the independent and identically distributed (i.i.d.) Bernoulli dropout variable with success probability *p*. The data are split into training and validation sets. The validation split percentage, dropout ratio, and number of neurons in the first, second, and third hidden layers are optimized using the Bayesian optimization algorithm described in the following section.

4.2.2.2 Bayesian Optimization of Hidden Layers with Gaussian Process Priors

The Bayesian optimization aims to utilize Bayes' rule to find the minimum or maximum of an objective function *f(x)* within a bounded set χ. In comparison with the traditional grid search method, which requires a significant amount of computational

power, the Bayesian algorithm obtains an optimal set of solutions with less iterations, which is critical for online diagnostic applications. The Bayesian optimization is able to automatically quantify the uncertainty of the minimizer or maximizer. For a generic family of models with data observation *D* and parameter $x \in \chi$, we assume a prior distribution *p(x)* and a likelihood $p(D \mid x)$ with given data *D*. We can infer the posterior distribution using Bayes' rule $p(x \mid D) \propto p(D \mid x)p(x)$. The maximizer (or minimizer) of the objective function then follows the maximum a posteriori (MAP) probability:

$$x^* = \arg\min_{x \in \chi} f(x) = \arg\min_{x \in \chi} (x \mid D) \tag{172}$$

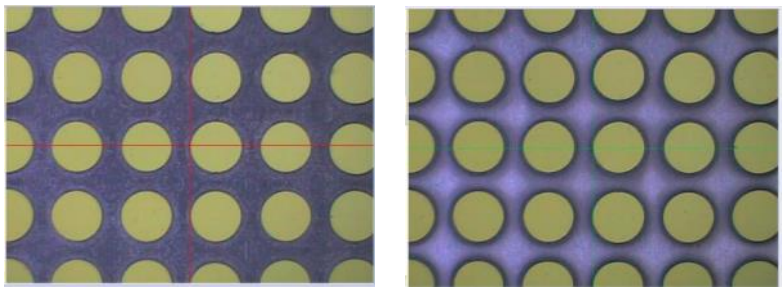
In this case, domain χ will be the range of the parameter within the network. We assume the Gaussian process prior such that the observation $D_{1:t} = \{x_{1:t,} y_{1:t}\}$ follows $y_{1:t} \sim Normal(f(x_{1:t}), \sum(x_{1:t,} x_{1:t}))$. Then the posterior probability distribution has the form $f(x) \mid f(x_{1:t}) \sim Normal(\mu_t(x \mid D), \sigma_t^2(x \mid D))$, where $\mu_t(x \mid D)$ and $\sigma_t^2(x \mid D)$ have complicated forms that are usually intractable. The Bayesian optimization establishes a probability model for *f(x)* by selecting various parameter values within the set χ. The model stores the previous calculated *f(x)* value and evaluation the area with higher probability to generate the minimum or maximum of *f(x)* without relying on the local gradient [192]. Rather than using an exhaustive search algorithm, the Bayesian optimization targets the area with a higher density of lower values of the cost function *f(x)*, which reduces the computation effort significantly. The basic algorithm [193] is described as follows:

We assumed an optimized solution $\hat{x}$ which solves the relationship $\hat{x} = \arg\min_x f(x) = \arg\min_{x \in \chi} p(f(x) \mid f(x_{1:t}))$ or $\hat{x} = \arg\max_x f(x) = \arg\max_{x \in \chi} p(f(x) \mid f(x_{1:t}))$, the objective is assumed to be a continuous function that is differentiable. For t = 1, 2 , … , select various x_t by optimizing the acquisition function over the Gaussian process: $x_t = \arg\max_x u(x \mid D_{1:t-1})$, where $D_{1:t}$ stores the previous observations $\{x_{1:t}, y_{1:t}\}$. In this case, a prior distribution *p(f(x))* can be assumed as normal distribution. The dataset $D_{1:t}$ will be augmented with the newly acquired $\{x_t, y_t\}$. Then, the whole process will repeat until certain criteria are met. The Bayesian optimization algorithm is combined with the CNN deep learning network to optimize the parameters within the network while reducing the training MSE. The Bayesian optimization is implemented to automatically tune the parameters in the CNN, playing the similar role as the widely used cross-validation. The next section explains the case study with the ECM drilling process.

4.2.3 Experimental Study

The experimental data collected in [3, 264, 274] were adopted to validate the proposed deep convolutional network. The electrolyte velocity at the outlet of the pump was set at 10 m/s; the average electrolyte temperature was measured to be 27 °C; the initial gap between the tool and the workpiece was set at 100 µm; the total tool travel was 800 µm; the workpiece material was 304 stainless steel; the electrolyte used was 10% wt. $NaNO_3$; the nominal diameter of the hole to be drilled was 900 µm; and the depth of the hole was 500 µm. The voltage, pulse-on-time, and feed rate were used as the controllable

process parameters because of the relatively low difficulty in measuring these factors, while the entry diameter of the micro-hole D_{in} and the exit diameter D_{out} were the response variables. Figure 108 shows the charge coupled device (CCD) camera image of the array of holes drilled during the μ-ECM experiment. It was observed that even the cutting parameters such as feed rate, pulse-on time, and voltage were the same. The patterns of the drilled hole were different. The possible causes were determined from the difference in electrolyte velocity at different parts of the workpiece, the interaction of the electrical fields between tools, and the existing active ions at different areas of the workpiece.



(**a**) (**b**)

Figure 108 – Picture taken by charge coupled device (CCD) camera. (a) The entry side of the hole (D_{in}); (b) the exit side of the hole (D_{out}) [3]

In order to create a forward prediction model for ECM drilling, the design of the experiment was implemented with three different sets of experiments generated. The cutting input parameter and output parameter were recorded. The experimental data results are shown in Table 19 in Appendix A.

4.2.4 Results

In this section, a new approach using the deep CNN with Bayesian optimization was introduced to increase the prediction accuracy and rate of convergence for the ECM drilling process. Rather than tuning the parameter by hand or using the grid and random search method, the Bayesian-based optimization algorithm helps navigate to the optimal set of parameters of the CNN. An initial guess of the network structure is made by trial and error to search for the network structure. Then the boundary of the parameter to be optimized is determined based on knowledge of the process and previous documented research. The Bayesian algorithm searches within the prescribed boundary and finds areas that have a higher probability of obtaining the optimal set of parameters. With sufficient sampling, an optimal set of parameters of the deep convolutional network is obtained. The automatic parameter tuning does not require knowledge of the ECM process and the network structure, which is ideal for industrial applications. The deep convolutional network prediction model is compared with a traditional neural network and physics-based NN. The proposed model has the advantages of requiring fewer training iterations to converge and fewer prediction errors in comparison with the previously proposed prediction models. In addition, with the added random dropout layer, the possibility of overfitting the network is decreased. With enough experimental data and more inputs, the proposed model offers a viable route to predict the relationship among the input and output parameters of an ECM process. The algorithm can be built into the computer system of an ECM drilling machine to facilitate the accurate process control and improve the throughput of the ECM drilling process.

Appendix A

Table 19 – Data used for training and testing the neural network [3]

No.	Voltage (V)	Pulse-On Time (µs)	Feed Rate (µm/s)	Din (µm)	Dout (µm)	Taper	Overcut (µm)
1	16	25	8	893	860	0.066	3.5
2	18	25	8	929	913	0.032	14.5
3	20	25	8	923	910	0.026	11.5
4	16	25	6	904	892	0.024	2
5	18	25	6	934	931	0.006	17
6	20	25	6	999	977	0.044	49.5
7	16	25	4	983	979	0.008	41.5
8	18	25	4	1050	1045	0.01	75
9	20	25	4	1125	1123	0.004	112.5

10	8	50	8	657.5	627.5	0.06	121.25
11	10	50	8	809.5	807.25	0.0045	45.25
12	12	50	8	866.25	858	0.0165	16.875
13	8	50	6	760	741	0.038	70
14	10	50	6	828.5	829.5	0.002	35.75
15	12	50	6	908.75	905.5	0.0065	4.375
16	8	50	4	781.75	780.25	0.003	59.125
17	10	50	4	887.25	881.75	0.011	6.375
18	12	50	4	957.75	970	0.0245	28.875
19	8	60	8	771.33	759.33	0.024	64.335
20	10	60	8	806.75	799.5	0.0145	46.625
21	12	60	8	862.75	847	0.0315	18.625
22	8	60	6	756.5	739.75	0.0335	71.75

23	10	60	6	776.75	777.5	0.0015	61.625
24	12	60	6	840.25	841.25	0.002	29.875
25	8	60	4	769	771.5	0.005	65.5
26	10	60	4	854.75	865.25	0.021	22.625
27	12	60	4	928.25	945.5	0.0345	14.125
28	8	70	8	718	721.5	0.007	91
29	10	70	8	779	796.75	0.0355	60.5
30	12	70	8	841.5	849.75	0.0165	29.25
31	8	70	6	736.5	744.5	0.016	81.75
32	10	70	6	802	829.75	0.0555	49
33	12	70	6	858.75	865	0.0125	20.625
34	8	70	4	783.25	783.25	0	58.375
35	10	70	4	878.75	872	0.0135	10.625

36	12	70	4	946.25	955.25	0.018	23.125
37	8	50	8	874	704	0.34	13
38	9	50	8	914	789	0.25	7
39	10	50	8	999	827	0.344	49.5
40	8	50	6	922	765	0.314	11
41	9	50	6	955	807	0.296	27.5
42	10	50	6	1039	837	0.404	69.5
43	8	50	4	932	797	0.27	16
44	9	50	4	1044	790	0.508	72
45	10	50	4	1130	858	0.544	115
46	8	60	8	903	708	0.39	1.5
47	9	60	8	967	766	0.402	33.5
48	10	60	8	1084	817	0.534	92

49	8	60	6	917	760	0.314	8.5
50	9	60	6	1043	856	0.374	71.5
51	10	60	6	1115	871	0.488	107.5
52	8	60	4	1071	754	0.634	85.5
53	9	60	4	1087	972	0.23	93.5
54	10	60	4	1263	1044	0.438	181.5
55	8	70	8	875	789	0.172	12.5
56	9	70	8	1071	842	0.458	85.5
57	10	70	8	1158	862	0.592	129
58	8	70	6	987	846	0.282	43.5
59	9	70	6	1212	886	0.652	156
60	10	70	6	1243	1056	0.374	171.5
61	8	70	4	1134	877	0.514	117

62	9	70	4	1260	935	0.65	180
63	10	70	4	1348	1016	0.664	224

CONCLUSION AND FUTURE WORK

This first part of this study examined several diagnostic and prognostic approaches implemented for the condition monitoring of rotating machineries and various parameter optimization algorithms to improve model accuracy and facilitate parameter selection without human intervention. The alternating parameter tuning algorithm of the complex wavelet can help improve the signal to noise ratio for the fault signal to a certain degree. However, the result is not ideal to be used in actual applications. The effect of each parameter to the signal to noise ratio remains unclear. A more robust parameter tuning scheme such as the Bayesian Optimization method should be attempted in the future study. In comparison with the complex wavelet, the empirical wavelet is a more promising techniques to extract fault signatures. The wavelet form is closer to the vibration pattern of the bearing signals. Hence, better extraction of features can be achieved. The parameters and shape of the wavelet implemented have not been thoroughly examined. A possible future study could involve examinations of the wavelet shape function and the parameters used in the EWT.

In comparison with the wavelet family, the dictionary learning algorithm is more robust and easier to implement. This research study focused on development of fast parameter updating algorithm of dictionary learning. An adaptive dictionary learning algorithm in the change of weights of the dictionaries is developed to be used in bearing fault diagnosis. The WLS and UKF can be implemented to update the dictionaries in order to accommodate various degradation stages of rolling element bearings. The WLS is easy to implement and yield satisfactory result. The UKF requires an artificially created variable

to be served as an internal variable in order to improve the diagnosis accuracy. It is also found that the performance of the UKF algorithm could be unstable. The developed methods focus on updating of model parameters. In the future study, structural change is should be examined instead of using an exhaustive search method. The challenge lies in the development of structural adaptation algorithm. The updated dictionary weights could be a good starting point to research the structural adaptation. Theoretically, a succinct subset of the dictionaries can be obtained based on the assigned weight value of each dictionary.

The third family of the diagnostic approaches developed belong to the classification family, tight clustering and deep learning. Both of these two methods use the CEEMD as a preprocessing tool. The CEEMD requires an added white noise to cancel the noise existed in the acquired signal. The standard deviation of the added white noise is obtained through the bootstrap resampling method. After the processing of the vibration signal and extraction of the critical IMFs. The CEEMD processed signal shows a clear trend of the bearing degradation. Fault size and degradation stages are calculated and classified using the tight clustering method and the deep convolutional neural network. The Bayesian optimization method is implemented to optimize the CNN structure and parameters. Both methods yield correct classification accuracy above 90% in comparison to the traditionally used clustering and SVM method which are around 70-80% in accuracy. One of the problems with the Bayesian optimization is the possibility of convergence on local extrema. In the future study, the selection of the range of parameters should be examined. Additionally, although the CEEMD successfully extracts desirable fault signature, the processing time of the CEEMD is relatively long, approximately a few hours to fully decompose the signal. More

computation efficient algorithm could be developed in the future to further facilitate the online application of condition monitoring.

The second part of the study focus on the prognosis of bearing life. The ARMA model is implemented with VFF-RLS to predict bearing degradation trend. With the VFF added into the traditional RLS, the converging speed of the RLS improves significantly. However, this method is only limited to short term prediction of bearing life. In comparison with the ARMA+VFF-RLS algorithm, the WNN model can predict the long-term behavior of the bearings based on load, and features from the vibration signal. Significant improvement for bearing prognosis approach should be expected in the future. As the current method only yield 70-80% accuracy. The lack of correlation between the vibration signal and the defect size prevents the added physics into the intelligent models. Perhaps the deep neural network and the multi-output least square support vector regression could be much beneficial to be implemented for bearing prognosis.

The last part of the study expands the developed algorithm into other areas. The ECM drilling process data is used to validate the physics based and deep neural network prediction capability. The deep learning algorithm and the MLS-SVM algorithm are both promising tools to map complex relationship between input and output process parameters. However, the training data size determines the accuracy of the prediction model. Because the limitation of obtaining experimental data, the proposed algorithm need to be further examined in the future. An interesting area to be studies could involve implementation of the deep neural network in prediction the cutting pattern of a standard ECM electrode. This could lead to significant expansion of the ECM process. Because with controlling of the cutting toolpath and capability of knowing the cutting shape of the tool, various geometry

could be generated rather than making dedicated electrode. The most efficient machining toolpath may also be an area to research. For instance, the determination of maximal material removal rate by maximizing the feed rate and reducing the depth of cut or by maximizing the depth of cut and reducing feed rate. No documented research has proposed a systematic way to decide the optimized tool path. The outcome will be critical in the future ECM implementation.

REFERENCES

1. Randall, R.B., *Vibration-based condition monitoring: industrial, aerospace and automotive applications*. 2011: John Wiley & Sons.

2. Braun, S., *Mechanical signature analysis: theory and applications.* 1986.

3. Rajora, M., et al., *A split-optimization approach for obtaining multiple solutions in single-objective process parameter optimization.* SpringerPlus, 2016. **5**(1): p. 1424.

4. Graney, B.P. and K. Starry, *Rolling element bearing analysis.* Materials Evaluation, 2012. **70**(1).

5. Randall, R.B. and J. Antoni, *Rolling element bearing diagnostics—A tutorial.* Mechanical systems and signal processing, 2011. **25**(2): p. 485-520.

6. Kurfess, T.R., S. Billington, and S.Y. Liang, *Advanced diagnostic and prognostic techniques for rolling element bearings*, in *Condition monitoring and control for intelligent manufacturing*. 2006, Springer. p. 137-165.

7. Li, Y., *Dynamic prognostics of rolling element bearing condition*. 1999, Georgia Institute of Technology.

8. Rai, A. and S. Upadhyay, *A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings.* Tribology International, 2016. **96**: p. 289-306.

9. Rai, V. and A. Mohanty, *Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert–Huang transform.* Mechanical Systems and Signal Processing, 2007. **21**(6): p. 2607-2615.

10. Peng, Z., W.T. Peter, and F. Chu, *A comparison study of improved Hilbert–Huang transform and wavelet transform: application to fault diagnosis for rolling bearing.* Mechanical systems and signal processing, 2005. **19**(5): p. 974-988.

11. Kim, Y.-H., B.D. Lim, and W.S. Cheoung, *Fault detection in a ball bearing system using a moving window.* Mechanical systems and signal processing, 1991. **5**(6): p. 461-473.

12. Schmidt, S., P.S. Heyns, and J.P. De Villiers, *A tacholess order tracking methodology based on a probabilistic approach to incorporate angular acceleration information into the maxima tracking process.* Mechanical Systems and Signal Processing, 2018. **100**: p. 630-646.

13. Zhao, M., et al., *A tacho-less order tracking technique for large speed variations.* Mechanical Systems and Signal Processing, 2013. **40**(1): p. 76-90.

14. Antoni, J., *The infogram: Entropic evidence of the signature of repetitive transients.* Mechanical Systems and Signal Processing, 2016. **74**: p. 73-94.

15. Mallat, S.G., *A theory for multiresolution signal decomposition: the wavelet representation.* IEEE transactions on pattern analysis and machine intelligence, 1989. **11**(7): p. 674-693.

16. Qiu, H., et al., *Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics.* Journal of sound and vibration, 2006. **289**(4-5): p. 1066-1090.

17. Rubini, R. and U. Meneghetti, *Application of the envelope and wavelet transform analyses for the diagnosis of incipient faults in ball bearings.* Mechanical systems and signal processing, 2001. **15**(2): p. 287-302.

18. Prabhakar, S., A. Mohanty, and A. Sekhar, *Application of discrete wavelet transform for detection of ball bearing race faults.* Tribology International, 2002. **35**(12): p. 793-800.

19. Selesnick, I.W., R.G. Baraniuk, and N.C. Kingsbury, *The dual-tree complex wavelet transform.* IEEE signal processing magazine, 2005. **22**(6): p. 123-151.

20. Peter, W.T., W.-x. Yang, and H. Tam, *Machine fault diagnosis through an effective exact wavelet analysis.* Journal of sound and vibration, 2004. **277**(4-5): p. 1005-1024.

21. Starck, J.-L., M. Elad, and D.L. Donoho, *Image decomposition via the combination of sparse representations and a variational approach.* IEEE transactions on image processing, 2005. **14**(10): p. 1570-1582.

22. Chaibi, S., et al., *A reliable approach to distinguish between transient with and without HFOs using TQWT and MCA.* Journal of neuroscience methods, 2014. **232**: p. 36-46.

23. Lajnef, T., et al., *Sleep spindle and K-complex detection using tunable Q-factor wavelet transform and morphological component analysis.* Frontiers in human neuroscience, 2015. **9**: p. 414.

24. Imani, E., M. Javidi, and H.-R. Pourreza, *Improvement of retinal blood vessel detection using morphological component analysis.* Computer methods and programs in biomedicine, 2015. **118**(3): p. 263-279.

25. Jolliffe, I.T., *Principal component analysis and factor analysis*, in *Principal component analysis*. 1986, Springer. p. 115-128.

26. Li, Y., et al., *Adaptive prognostics for rolling element bearing condition.* Mechanical systems and signal processing, 1999. **13**(1): p. 103-113.

27. Khanam, S., N. Tandon, and J. Dutt, *Fault size estimation in the outer race of ball bearing using discrete wavelet transform of the vibration signal.* Procedia Technology, 2014. **14**: p. 12-19.

28. Rao, V.V. and C. Ratnam, *Estimation of Defect Severity in Rolling Element Bearings using Vibration Signals with Artificial Neural Network.* Jordan Journal of Mechanical & Industrial Engineering, 2015. **9**(2).

29. Mallat, S., *A wavelet tour of signal processing*. 1999: Academic press.

30. Choi, H., et al. *Hidden Markov tree modeling of complex wavelet transforms*. in *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*. 2000. IEEE.

31. Dragotti, P.L. and M. Vetterli, *Wavelet footprints: theory, algorithms, and applications.* IEEE Transactions on Signal Processing, 2003. **51**(5): p. 1306-1323.

32. Abdi, H. and L.J. Williams, *Principal component analysis.* Wiley interdisciplinary reviews: computational statistics, 2010. **2**(4): p. 433-459.

33. Starck, J.-L., D. Donoho, and M. Elad, *Redundant multiscale transforms and their application for morphological component separation*. 2004, CM-P00052061.

34. Ge, D., X. Jiang, and Y. Ye, *A note on the complexity of Lp minimization.* Mathematical programming, 2011. **129**(2): p. 285-299.

35. Sohaib, M., C.-H. Kim, and J.-M. Kim, *A Hybrid Feature Model and Deep-Learning-Based Bearing Fault Diagnosis.* Sensors, 2017. **17**(12): p. 2876.

36. He, W., et al., *Tunable Q-factor wavelet transform denoising with neighboring coefficients and its application to rotating machinery fault diagnosis.* Science China Technological Sciences, 2013. **56**(8): p. 1956-1965.

37. He, W. and Y. Zi. *Sparsity-assisted signal representation for rotating machinery fault diagnosis using the tunable Q-factor wavelet transform with overlapping group shrinkage*. in *Wavelet Analysis and Pattern Recognition (ICWAPR), 2014 International Conference on*. 2014. IEEE.

38. Selesnick, I.W. *Sparse signal representations using the tunable Q-factor wavelet transform*. in *Wavelets and Sparsity XIV*. 2011. International Society for Optics and Photonics.

39. Selesnick, I.W., *Wavelet transform with tunable Q-factor.* IEEE transactions on signal processing, 2011. **59**(8): p. 3560-3575.

40. Selesnick, I.W., *Resonance-based signal decomposition: A new sparsity-enabled signal analysis method.* Signal Processing, 2011. **91**(12): p. 2793-2809.

41. Li, Q. and S.Y. Liang, *Incipient multi-fault diagnosis of rolling bearing using improved TQWT and sparse representation approach*, in *2017 IEEE 2nd International Conference on Signal and Image Processing (ICSIP)*. 2017, IEEE: Singapore, Singapore.

42. Lu, Y., Q. Li, and S.Y. Liang. *Adaptive prognosis of bearing degradation based on wavelet decomposition assisted ARMA model*. in *Technology, Networking, Electronic and Automation Control Conference (ITNEC), 2017 IEEE 2nd Information*. 2017. IEEE.

43. Lu, Y., et al., *Prognosis of Bearing Degradation using Gradient Variable Forgetting Factor RLS Combined with Time Series Model.* IEEE Access, 2018.

44. Luo, H., et al., *Synthesized synchronous sampling technique for differential bearing damage detection.* Journal of Engineering for Gas Turbines and Power, 2010. **132**(7): p. 072501.

45. Siegel, D., et al., *Novel method for rolling element bearing health assessment—A tachometer-less synchronously averaged envelope feature extraction technique.* Mechanical Systems and Signal Processing, 2012. **29**: p. 362-376.

46. Wang, Y., et al., *An online tacholess order tracking technique based on generalized demodulation for rolling bearing fault detection.* Journal of sound and vibration, 2016. **367**: p. 233-249.

47. Feng, Z., X. Chen, and T. Wang, *Time-varying demodulation analysis for rolling bearing fault diagnosis under variable speed conditions.* Journal of Sound and Vibration, 2017. **400**: p. 71-85.

48. McFadden, P. and J. Smith, *Vibration monitoring of rolling element bearings by the high-frequency resonance technique—a review.* Tribology international, 1984. **17**(1): p. 3-10.

49. Randall, R.B., J. Antoni, and S. Chobsaard, *The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals.* Mechanical systems and signal processing, 2001. **15**(5): p. 945-962.

50. Antoni, J., *The spectral kurtosis: a useful tool for characterising non-stationary signals.* Mechanical Systems and Signal Processing, 2006. **20**(2): p. 282-307.

51. Antoni, J., *Fast computation of the kurtogram for the detection of transient faults.* Mechanical Systems and Signal Processing, 2007. **21**(1): p. 108-124.

52. Antoni, J. and R. Randall, *The spectral kurtosis: application to the vibratory surveillance and diagnostics of rotating machines.* Mechanical systems and signal processing, 2006. **20**(2): p. 308-331.

53. Huang, N.E., et al. *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*. in *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences*. 1998. The Royal Society.

54. Lei, Y., et al., *A review on empirical mode decomposition in fault diagnosis of rotating machinery.* Mechanical Systems and Signal Processing, 2013. **35**(1-2): p. 108-126.

55. Daubechies, I., J. Lu, and H.-T. Wu, *Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool.* Applied and computational harmonic analysis, 2011. **30**(2): p. 243-261.

56. Torres, M.E., et al. *A complete ensemble empirical mode decomposition with adaptive noise*. in *Acoustics, speech and signal processing (ICASSP), 2011 IEEE international conference on*. 2011. IEEE.

57. Gilles, J., *Empirical wavelet transform.* IEEE transactions on signal processing, 2013. **61**(16): p. 3999-4010.

58. Chen, J., et al., *Generator bearing fault diagnosis for wind turbine via empirical wavelet transform using measured vibration signals.* Renewable Energy, 2016. **89**: p. 80-92.

59. Borghesani, P., et al., *A new procedure for using envelope analysis for rolling element bearing diagnostics in variable operating conditions.* Mechanical Systems and Signal Processing, 2013. **38**(1): p. 23-35.

60. Antoni, J., *Cyclic spectral analysis of rolling-element bearing signals: Facts and fictions.* Journal of Sound and vibration, 2007. **304**(3-5): p. 497-529.

61. Ali, J.B., et al., *Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals.* Applied Acoustics, 2015. **89**: p. 16-27.

62. Dybała, J. and R. Zimroz, *Rolling bearing diagnosing method based on empirical mode decomposition of machine vibration signal.* Applied Acoustics, 2014. **77**: p. 195-203.

63. Zhang, X. and J. Zhou, *Multi-fault diagnosis for rolling element bearings based on ensemble empirical mode decomposition and optimized support vector machines.* Mechanical Systems and Signal Processing, 2013. **41**(1-2): p. 127-140.

64. Zhang, J., et al., *Performance enhancement of ensemble empirical mode decomposition.* Mechanical Systems and Signal Processing, 2010. **24**(7): p. 2104-2123.

65. Lei, Y., Z. He, and Y. Zi, *Application of the EEMD method to rotor fault diagnosis of rotating machinery.* Mechanical Systems and Signal Processing, 2009. **23**(4): p. 1327-1338.

66. Kedadouche, M., M. Thomas, and A. Tahan, *A comparative study between Empirical Wavelet Transforms and Empirical Mode Decomposition Methods: Application to bearing defect diagnosis.* Mechanical Systems and Signal Processing, 2016. **81**: p. 88-107.

67. Daubechies, I., *Ten lectures on wavelets*. Vol. 61. 1992: Siam.

68. Chen, S.S., D.L. Donoho, and M.A. Saunders, *Atomic decomposition by basis pursuit.* SIAM review, 2001. **43**(1): p. 129-159.

69. Tibshirani, R., *Regression shrinkage and selection via the lasso.* Journal of the Royal Statistical Society. Series B (Methodological), 1996: p. 267-288.

70. Li, F., et al. *Optimal Lq norm regularization for sparse reflectivity inversion*. in *2017 SEG International Exposition and Annual Meeting*. 2017. Society of Exploration Geophysicists.

71. Marjanovic, G. and V. Solo, *On lq optimization and matrix completion.* IEEE Transactions on signal processing, 2012. **60**(11): p. 5714-5724.

72. Raskutti, G., M.J. Wainwright, and B. Yu, *Minimax rates of estimation for high-dimensional linear regression over lq balls.* IEEE transactions on information theory, 2011. **57**(10): p. 6976-6994.

73. Liang, S.Y., et al., *Adaptive prognostics for rotary machineries.* Procedia Engineering, 2014. **86**: p. 852-857.

74. Igba, J., et al., *Analysing RMS and peak values of vibration signals for condition monitoring of wind turbine gearboxes.* Renewable Energy, 2016. **91**: p. 90-106.

75. Lu, Y., et al., *Prognosis of Bearing Degradation Using Gradient Variable Forgetting Factor RLS Combined With Time Series Model.* IEEE Access, 2018. **6**: p. 10986-10995.

76. Samanta, B. and K. Al-Balushi, *Artificial neural network based fault diagnostics of rolling element bearings using time-domain features.* Mechanical systems and signal processing, 2003. **17**(2): p. 317-328.

77. Bonnardot, F., R. Randall, and J. Antoni, *Enhanced unsupervised noise cancellation using angular resampling for planetary bearing fault diagnosis.*

78. Sawalhi, N. and R. Randall, *Helicopter gearbox bearing blind fault identification using a range of analysis techniques.* Australian Journal of Mechanical Engineering, 2008. **5**(2): p. 157-168.

79. Bai, M., et al., *Fault diagnosis of rotating machinery using an intelligent order tracking system.* Journal of sound and vibration, 2005. **280**(3-5): p. 699-718.

80. Wang, J., Y. Peng, and W. Qiao, *Current-aided order tracking of vibration signals for bearing fault diagnosis of direct-drive wind turbines.* IEEE Transactions on Industrial Electronics, 2016. **63**(10): p. 6336-6346.

81. Wang, Y., et al., *Tacholess order-tracking approach for wind turbine gearbox fault detection.* Frontiers of Mechanical Engineering, 2017. **12**(3): p. 427-439.

82. Zhao, M., et al., *Tacholess envelope order analysis and its application to fault detection of rolling element bearings with varying speeds.* Sensors, 2013. **13**(8): p. 10856-10875.

83. Stack, J.R., R.G. Harley, and T.G. Habetler, *An amplitude modulation detector for fault diagnosis in rolling element bearings.* IEEE Transactions on Industrial Electronics, 2004. **51**(5): p. 1097-1102.

84. Gong, X. and W. Qiao, *Bearing fault diagnosis for direct-drive wind turbines via current-demodulated signals.* IEEE Transactions on Industrial Electronics, 2013. **60**(8): p. 3419-3428.

85. Griffin, D. and J. Lim, *Signal estimation from modified short-time Fourier transform.* IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984. **32**(2): p. 236-243.

86. Lei, Y., et al., *Application of an improved kurtogram method for fault diagnosis of rolling element bearings.* Mechanical Systems and Signal Processing, 2011. **25**(5): p. 1738-1749.

87. Segreto, T., et al., *Vibration Sensor Monitoring of Nickel-Titanium Alloy Turning for Machinability Evaluation.* Sensors, 2017. **17**(12): p. 2885.

88. Yan, R., R.X. Gao, and X. Chen, *Wavelets for fault diagnosis of rotary machines: A review with applications.* Signal processing, 2014. **96**: p. 1-15.

89. Mairal, J., et al. *Online dictionary learning for sparse coding*. in *Proceedings of the 26th annual international conference on machine learning*. 2009. ACM.

90. Mairal, J., et al., *Online learning for matrix factorization and sparse coding.* Journal of Machine Learning Research, 2010. **11**(Jan): p. 19-60.

91. Aharon, M., M. Elad, and A. Bruckstein, *$ rm k $-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.* IEEE Transactions on signal processing, 2006. **54**(11): p. 4311-4322.

92. Rubinstein, R., T. Peleg, and M. Elad, *Analysis K-SVD: A dictionary-learning algorithm for the analysis sparse model.* IEEE Transactions on Signal Processing, 2013. **61**(3): p. 661-677.

93. Wang, L., et al., *IK-SVD: dictionary learning for spatial big data via incremental atom update.* Computing in Science & Engineering, 2014. **16**(4): p. 41-52.

94. Jiang, Z., Z. Lin, and L.S. Davis. *Learning a discriminative dictionary for sparse coding via label consistent K-SVD*. in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 2011. IEEE.

95. Tropp, J.A. and A.C. Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit.* IEEE Transactions on information theory, 2007. **53**(12): p. 4655-4666.

96. Donoho, D.L., et al., *Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit.* IEEE transactions on Information Theory, 2012. **58**(2): p. 1094-1121.

97. Rubinstein, R., M. Zibulevsky, and M. Elad, *Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit.* Cs Technion, 2008. **40**(8): p. 1-15.

98. Zhou, H., et al., *Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model.* Mechanical systems and signal processing, 2016. **72**: p. 65-79.

99. Yang, B., R. Liu, and X. Chen, *Fault diagnosis for a wind turbine generator bearing via sparse representation and shift-invariant K-SVD.* IEEE Transactions on Industrial Informatics, 2017. **13**(3): p. 1321-1331.

100. Golub, G.H. and C. Reinsch, *Singular value decomposition and least squares solutions.* Numerische mathematik, 1970. **14**(5): p. 403-420.

101. Nocedal, J. and S.J. Wright, *Conjugate gradient methods.* Numerical optimization, 2006: p. 101-134.

102. Daubechies, I., et al., *Iteratively reweighted least squares minimization for sparse recovery.* Communications on Pure and Applied Mathematics: A Journal Issued by the Courant Institute of Mathematical Sciences, 2010. **63**(1): p. 1-38.

103. Ayhan, B., C. Kwan, and S.Y. Liang. *Adaptive Remaining Useful Life Prediction Algorithm for Bearings*. in *2018 IEEE International Conference on Prognostics and Health Management (ICPHM)*. 2018. IEEE.

104. Smith, W.A. and R.B. Randall, *Rolling element bearing diagnostics using the Case Western Reserve University data: A benchmark study.* Mechanical Systems and Signal Processing, 2015. **64**: p. 100-131.

105. Tavner, P.J. and J. Penman, *Condition monitoring of electrical machines*. Vol. 1. 1987: John Wiley & Sons Incorporated.

106. Liu, R., et al., *Artificial intelligence for fault diagnosis of rotating machinery: A review*. Mechanical Systems and Signal Processing, 2018. **108**: p. 33-47.

107. Lu, Y., R. Xie, and S.Y. Liang, *Detection of weak fault using sparse empirical wavelet transform for cyclic fault.* The International Journal of Advanced Manufacturing Technology, 2018: p. 1-7.

108. Zhang, Y. and R. Randall, *Rolling element bearing fault diagnosis based on the combination of genetic algorithms and fast kurtogram.* Mechanical Systems and Signal Processing, 2009. **23**(5): p. 1509-1517.

109. Randall, R., N. Sawalhi, and M. Coats, *A comparison of methods for separation of deterministic and random signals.* International Journal of Condition Monitoring, 2011. **1**(1): p. 11-19.

110. Fyfe, K. and E. Munck, *Analysis of computed order tracking.* Mechanical Systems and Signal Processing, 1997. **11**(2): p. 187-205.

111. Bossley, K., et al., *Hybrid computed order tracking.* Mechanical systems and signal processing, 1999. **13**(4): p. 627-641.

112. Ahamed, N., Y. Pandya, and A. Parey, *Spur gear tooth root crack detection using time synchronous averaging under fluctuating speed.* Measurement, 2014. **52**: p. 1-11.

113. Combet, F. and L. Gelman, *An automated methodology for performing time synchronous averaging of a gearbox signal without speed sensor.* Mechanical systems and signal processing, 2007. **21**(6): p. 2590-2606.

114. Bonnardot, F., et al., *Use of the acceleration signal of a gearbox in order to perform angular resampling (with limited speed fluctuation).* Mechanical Systems and Signal Processing, 2005. **19**(4): p. 766-785.

115. Elasha, F., et al., *A comparative study of the effectiveness of adaptive filter algorithms, spectral kurtosis and linear prediction in detection of a naturally degraded bearing in a gearbox.* Journal of Failure Analysis and Prevention, 2014. **14**(5): p. 623-636.

116. Akaike, H., *Fitting autoregressive models for prediction.* Annals of the institute of Statistical Mathematics, 1969. **21**(1): p. 243-247.

117. Box, G.E., et al., *Time series analysis: forecasting and control*. 2015: John Wiley & Sons.

118. Daubechies, I., *The wavelet transform, time-frequency localization and signal analysis.* IEEE transactions on information theory, 1990. **36**(5): p. 961-1005.

119. Sun, Z. and C. Chang, *Structural damage assessment based on wavelet packet transform.* Journal of structural engineering, 2002. **128**(10): p. 1354-1361.

120. Wang, D. and K.-L. Tsui, *Dynamic Bayesian wavelet transform: New methodology for extraction of repetitive transients.* Mechanical Systems and Signal Processing, 2017. **88**: p. 137-144.

121. Zhang, Q. and B. Li. *Discriminative K-SVD for dictionary learning in face recognition*. in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. 2010. IEEE.

122. Aharon, M., M. Elad, and A. Bruckstein, *K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation.* IEEE Transactions on signal processing, 2006. **54**(11): p. 4311.

123. Skretting, K. and K. Engan. *Image compression using learned dictionaries by RLS-DLA and compared with K-SVD*. in *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. 2011. IEEE.

124. Gharavi-Alkhansari, M. and T.S. Huang. *A fast orthogonal matching pursuit algorithm*. in *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*. 1998. IEEE.

125. Borghesani, P., et al., *Testing second order cyclostationarity in the squared envelope spectrum of non-white vibration signals.* Mechanical Systems and Signal Processing, 2013. **40**(1): p. 38-55.

126. Feng, Z., H. Ma, and M.J. Zuo, *Spectral negentropy based sidebands and demodulation analysis for planet bearing fault diagnosis.* Journal of Sound and Vibration, 2017. **410**: p. 124-150.

127. Julier, S.J. and J.K. Uhlmann. *New extension of the Kalman filter to nonlinear systems*. in *Signal processing, sensor fusion, and target recognition VI*. 1997. International Society for Optics and Photonics.

128. Wan, E.A. and R. Van Der Merwe, *The unscented Kalman filter*. Kalman filtering and neural networks, 2001: p. 221-280.

129. Julier, S.J. and J.K. Uhlmann, *Unscented filtering and nonlinear estimation.* Proceedings of the IEEE, 2004. **92**(3): p. 401-422.

130. Box, G.E. and G.C. Tiao, *Bayesian inference in statistical analysis*. Vol. 40. 2011: John Wiley & Sons.

131. Lu, Y., R. Xie, and S.Y. Liang, *Adaptive online dictionary learning for bearing fault diagnosis.* The International Journal of Advanced Manufacturing Technology, 2018: p. 1-8.

132. Yang, Y., et al., *Application of a specialized capacitance probe in bearing diagnosis.* Wear, 1999. **225**: p. 1215-1221.

133. Abboud, D., et al., *Advanced bearing diagnostics: A comparative study of two powerful approaches.* Mechanical Systems and Signal Processing, 2019. **114**: p. 604-627.

134. Lu, Y., R. Xie, and S.Y. Liang, *Detection of weak fault using sparse empirical wavelet transform for cyclic fault.* The International Journal of Advanced Manufacturing Technology, 2018. **99**(5-8): p. 1195-1201.

135. Widodo, A. and B.-S. Yang, *Support vector machine in machine condition monitoring and fault diagnosis.* Mechanical systems and signal processing, 2007. **21**(6): p. 2560-2574.

136. Kim, H.-E., et al., *Bearing fault prognosis based on health state probability estimation.* Expert Systems with Applications, 2012. **39**(5): p. 5200-5213.

137. Jardine, A.K., D. Lin, and D. Banjevic, *A review on machinery diagnostics and prognostics implementing condition-based maintenance.* Mechanical systems and signal processing, 2006. **20**(7): p. 1483-1510.

138. Peng, Y., M. Dong, and M.J. Zuo, *Current status of machine prognostics in condition-based maintenance: a review.* The International Journal of Advanced Manufacturing Technology, 2010. **50**(1-4): p. 297-313.

139. Schoen, R.R., et al., *Motor bearing damage detection using stator current monitoring.* IEEE transactions on industry applications, 1995. **31**(6): p. 1274-1279.

140. Al-Ghamd, A.M. and D. Mba, *A comparative experimental study on the use of acoustic emission and vibration analysis for bearing defect identification and estimation of defect size.* Mechanical systems and signal processing, 2006. **20**(7): p. 1537-1571.

141. Shiroishi, J., et al., *Bearing condition diagnostics via vibration and acoustic emission measurements.* Mechanical systems and signal processing, 1997. **11**(5): p. 693-705.

142. Zhou, W., T.G. Habetler, and R.G. Harley. *Bearing condition monitoring methods for electric machines: A general review*. in *Diagnostics for Electric Machines, Power Electronics and Drives, 2007. SDEMPED 2007. IEEE International Symposium on*. 2007. IEEE.

143. Loutas, T., et al., *The combined use of vibration, acoustic emission and oil debris on-line monitoring towards a more effective condition monitoring of rotating machinery.* Mechanical systems and signal processing, 2011. **25**(4): p. 1339-1352.

144. Baillie, D. and J. Mathew, *A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings.* Mechanical Systems and Signal Processing, 1996. **10**(1): p. 1-17.

145. Zhang, X., et al. *An integrated approach to bearing fault diagnostics and prognostics*. in *American Control Conference, 2005. Proceedings of the 2005*. 2005. IEEE.

146. Mishra, C., A. Samantaray, and G. Chakraborty, *Rolling element bearing defect diagnosis under variable speed operation through angle synchronous averaging of wavelet de-noised estimate.* Mechanical Systems and Signal Processing, 2016. **72**: p. 206-222.

147. Yazici, B. and G.B. Kliman, *An adaptive statistical time-frequency method for detection of broken bars and bearing faults in motors using stator current.* IEEE Transactions on Industry Applications, 1999. **35**(2): p. 442-452.

148. Huang, N.E., et al., *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis.* Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences, 1998. **454**(1971): p. 903-995.

149. Yeh, J.-R., J.-S. Shieh, and N.E. Huang, *Complementary ensemble empirical mode decomposition: A novel noise enhanced data analysis method.* Advances in adaptive data analysis, 2010. **2**(02): p. 135-156.

150. Jain, A.K., M.N. Murty, and P.J. Flynn, *Data clustering: a review.* ACM computing surveys (CSUR), 1999. **31**(3): p. 264-323.

151. Yiakopoulos, C., K.C. Gryllias, and I.A. Antoniadis, *Rolling element bearing fault detection in industrial environments based on a K-means clustering approach.* Expert Systems with Applications, 2011. **38**(3): p. 2888-2911.

152. Yuwono, M., et al., *Automatic bearing fault diagnosis using particle swarm clustering and Hidden Markov Model.* Engineering Applications of Artificial Intelligence, 2016. **47**: p. 88-100.

153. Lei, Y., et al., *New clustering algorithm-based fault diagnosis using compensation distance evaluation technique.* Mechanical Systems and Signal Processing, 2008. **22**(2): p. 419-435.

154. Wentao, S., L. Changhou, and Z. Dan. *Bearing fault diagnosis based on feature weighted FCM cluster analysis*. in *Computer Science and Software Engineering, 2008 International Conference on*. 2008. IEEE.

155. Soualhi, A., G. Clerc, and H. Razik, *Detection and diagnosis of faults in induction motor using an improved artificial ant clustering technique.* IEEE Transactions on Industrial Electronics, 2013. **60**(9): p. 4053-4062.

156. Tseng, G.C. and W.H. Wong, *Tight clustering: a resampling‐based approach for identifying stable and tight patterns in data.* Biometrics, 2005. **61**(1): p. 10-16.

157. Allison, D.B., et al., *Microarray data analysis: from disarray to consolidation and consensus.* Nature reviews genetics, 2006. **7**(1): p. 55.

158. Maitra, R. and V. Melnykov, *Simulating data to study performance of finite mixture modeling and clustering algorithms.* Journal of Computational and Graphical Statistics, 2010. **19**(2): p. 354-376.

159. Maitra, R. and V. Melnykov, *SUPPLEMENT TO "ASSESSING SIGNIFICANCE IN FINITE MIXTURE MODELS".* Some theoretical contributions to the evaluation and assessment of finite mixture models with applications, 2009: p. 109.

160. Fraley, C. and A.E. Raftery, *How many clusters? Which clustering method? Answers via model-based cluster analysis.* The computer journal, 1998. **41**(8): p. 578-588.

161. Torres, M.E., et al. *A complete ensemble empirical mode decomposition with adaptive noise*. in *2011 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. 2011. IEEE.

162. Wu, Z. and N.E. Huang, *Ensemble empirical mode decomposition: a noise-assisted data analysis method.* Advances in adaptive data analysis, 2009. **1**(01): p. 1-41.

163. Jiang, L., J. Xuan, and T. Shi, *Feature extraction based on semi-supervised kernel Marginal Fisher analysis and its application in bearing fault diagnosis.* Mechanical Systems and Signal Processing, 2013. **41**(1-2): p. 113-126.

164. Chen, Z., et al., *Deep neural networks-based rolling bearing fault diagnosis.* Microelectronics Reliability, 2017. **75**: p. 327-333.

165. Tibshirani, R. and G. Walther, *Cluster validation by prediction strength.* Journal of Computational and Graphical Statistics, 2005. **14**(3): p. 511-528.

166. Renaudin, L., et al., *Natural roller bearing fault detection by angular measurement of true instantaneous angular speed.* Mechanical Systems and Signal Processing, 2010. **24**(7): p. 1998-2011.

167. Anderson, T.L., *Fracture mechanics: fundamentals and applications*. 2017: CRC press.

168. Borghesani, P., et al., *Application of cepstrum pre-whitening for the diagnosis of bearing faults under variable speed conditions.* Mechanical Systems and Signal Processing, 2013. **36**(2): p. 370-384.

169. Lu, Y., R. Xie, and S.Y. Liang, *Bearing fault diagnosis with nonlinear adaptive dictionary learning.* The International Journal of Advanced Manufacturing Technology: p. 1-13.

170. Mba, D. and R.B. Rao, *Development of Acoustic Emission Technology for Condition Monitoring and Diagnosis of Rotating Machines; Bearings, Pumps, Gearboxes, Engines and Rotating Structures.* 2006.

171. Bogue, R., *Sensors for condition monitoring: a review of technologies and applications.* Sensor Review, 2013. **33**(4): p. 295-299.

172. Bonnardot, F., et al., *Enhanced unsupervised noise cancellation using angular resampling for planetary bearing fault diagnosis.* International journal of acoustics and vibration, 2004. **9**(2): p. 51-60.

173. Lu, S., et al., *Fault diagnosis of motor bearing with speed fluctuation via angular resampling of transient sound signals.* Journal of Sound and Vibration, 2016. **385**: p. 16-32.

174. Bechhoefer, E. and M. Kingsley. *A review of time synchronous average algorithms*. in *Annual conference of the prognostics and health management society*. 2009.

175. Ur Rehman, N. and D.P. Mandic, *Filter bank property of multivariate empirical mode decomposition.* IEEE transactions on signal processing, 2011. **59**(5): p. 2421-2426.

176. Li, M., et al., *An improved method based on CEEMD for fault diagnosis of rolling bearing.* Advances in Mechanical Engineering, 2014. **6**: p. 676205.

177. Zhao, L., W. Yu, and R. Yan, *Rolling bearing fault diagnosis based on CEEMD and time series modeling.* Mathematical Problems in Engineering, 2014. **2014**.

178. Bian, J., et al. *Fault detection of rolling bearings through vibration analysis via the hybrid CEEMD-EMD approach*. in *2014 Prognostics and System Health Management Conference (PHM-2014 Hunan)*. 2014. IEEE.

179. Lu, Y., R. Xie, and S.Y. Liang, *CEEMD Assisted Bearing Degradation Assessment Using Tight Clustering.* The International Journal of Advanced Manufacturing Technology, 2019.

180. Mba, D., *The use of acoustic emission for estimation of bearing defect size.* Journal of Failure analysis and prevention, 2008. **8**(2): p. 188-192.

181. Al-Ghamdi, A.M., et al., *Estimation of bearing defect size with acoustic emission.* Insight-Non-Destructive Testing and Condition Monitoring, 2004. **46**(12): p. 758-761.

182. LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning.* nature, 2015. **521**(7553): p. 436.

183. Gan, M. and C. Wang, *Construction of hierarchical diagnosis network based on deep learning and its application in the fault pattern recognition of rolling element bearings.* Mechanical Systems and Signal Processing, 2016. **72**: p. 92-104.

184. Jia, F., et al., *Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data.* Mechanical Systems and Signal Processing, 2016. **72**: p. 303-315.

185. Guo, X., L. Chen, and C. Shen, *Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis.* Measurement, 2016. **93**: p. 490-502.

186. Tamilselvan, P. and P. Wang, *Failure diagnosis using deep belief learning based health state classification.* Reliability Engineering & System Safety, 2013. **115**: p. 124-135.

187. He, M. and D. He, *Deep learning based approach for bearing fault diagnosis.* IEEE Transactions on Industry Applications, 2017. **53**(3): p. 3057-3065.

188. Shao, H., et al., *A novel deep autoencoder feature learning method for rotating machinery fault diagnosis.* Mechanical Systems and Signal Processing, 2017. **95**: p. 187-204.

189. Radford, A., L. Metz, and S. Chintala, *Unsupervised representation learning with deep convolutional generative adversarial networks.* arXiv preprint arXiv:1511.06434, 2015.

190. Shin, H.-C., et al., *Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning.* IEEE transactions on medical imaging, 2016. **35**(5): p. 1285-1298.

191. Krizhevsky, A., I. Sutskever, and G.E. Hinton. *Imagenet classification with deep convolutional neural networks*. in *Advances in neural information processing systems*. 2012.

192. Snoek, J., H. Larochelle, and R.P. Adams. *Practical bayesian optimization of machine learning algorithms*. in *Advances in neural information processing systems*. 2012.

193. Brochu, E., V.M. Cora, and N. De Freitas, *A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning.* arXiv preprint arXiv:1012.2599, 2010.

194. Zhang, W., et al., *A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load.* Mechanical Systems and Signal Processing, 2018. **100**: p. 439-453.

195. Dyer, D. and R. Stewart, *Detection of rolling element bearing damage by statistical vibration analysis.* Journal of mechanical design, 1978. **100**(2): p. 229-235.

196. Randall, R., *Computer-aided Vibration Spectrum Trend Analysis for Condition Monitoring.* Maintenance Management International, 1985. **5**(3): p. 161-167.

197. Howard, I., *A Review of Rolling Element Bearing Vibration'Detection, Diagnosis and Prognosis'*. 1994, DEFENCE SCIENCE AND TECHNOLOGY ORGANIZATION CANBERRA (AUSTRALIA).

198. Paris, P. and F. Erdogan, *A critical analysis of crack propagation laws.* Journal of basic engineering, 1963. **85**(4): p. 528-533.

199. Yan, J., M. Koc, and J. Lee, *A prognostic algorithm for machine performance assessment and its application.* Production Planning & Control, 2004. **15**(8): p. 796-801.

200. Shao, Y. and K. Nezu, *Prognosis of remaining bearing life using neural networks.* Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering, 2000. **214**(3): p. 217-230.

201. Lundberg, G. and A. Palmgren, *Dynamic capacity of rolling bearings.* Journal of Applied Mechanics-Transactions of the ASME, 1949. **16**(2): p. 165-172.

202. Qiu, H., et al., *Wavelet filter-based weak signature detection method and its application on rolling element bearing prognostics.* Journal of sound and vibration, 2006. **289**(4): p. 1066-1090.

203. He, W., et al., *Automatic fault feature extraction of mechanical anomaly on induction motor bearing using ensemble super-wavelet transform.* Mechanical Systems and Signal Processing, 2015. **54**: p. 457-480.

204. Choi, B., *ARMA model identification*. 2012: Springer Science & Business Media.

205. Li, Q., et al., *Long Range Dependence Prognostics for Bearing Vibration Intensity Chaotic Time Series.* Entropy, 2016. **18**(1): p. 23.

206. Haykin, S.S., *Adaptive filter theory*. 2008: Pearson Education India.

207. Nectoux, P., et al. *PRONOSTIA: An experimental platform for bearings accelerated degradation tests*. in *IEEE International Conference on Prognostics and Health Management, PHM'12*. 2012. IEEE Catalog Number: CPF12PHM-CDR.

208. Lu, Y., Z. Pan, and S. Liang, *Bearing Prognosis using ARMA Model with Gradient Based Variable Forgetting Factor Recursive Least Square Algorithm* Frontiers of Information Technology & Electronic Engineering, 2017.

209. Gustafsson, O.G. and T. Tallian, *Detection of damage in assembled rolling element bearings.* ASLE TRANSACTIONS, 1962. **5**(1): p. 197-209.

210. Dyer, D. and R. Stewart, *Detection of rolling element bearing damage by statistical vibration analysis.* ASME J. Mech. Des, 1978. **100**(2): p. 229-235.

211. Alfredson, R. and J. Mathew, *Time domain methods for monitoring the condition of rolling element bearings.* NASA STI/Recon Technical Report A, 1985. **86**: p. 22166.

212. Martin, H. and F. Honarvar, *Application of statistical moments to bearing failure detection.* Applied acoustics, 1995. **44**(1): p. 67-77.

213. Zhou, W., et al., *Incipient bearing fault detection via motor stator current noise cancellation using wiener filter*. IEEE Transactions on Industry Applications, 2009. **45**(4): p. 1309-1317.

214. Tian, J., et al., *Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with K-nearest neighbor distance analysis.* IEEE Transactions on Industrial Electronics, 2016. **63**(3): p. 1793-1803.

215. Tian, J., C. Morillo, and M.G. Pecht. *Rolling element bearing fault diagnosis using simulated annealing optimized spectral kurtosis*. in *Prognostics and Health Management (PHM), 2013 IEEE Conference on*. 2013. IEEE.

216. Murakami, Y., M. Kaneta, and H. Yatsuzuka, *Analysis of surface crack propagation in lubricated rolling contact.* ASLE transactions, 1985. **28**(1): p. 60-68.

217. Bryant, M., *A pitting model for rolling contact fatigue.* 1983.

218. Pan, Z., et al., *Prediction of machining-induced phase transformation and grain growth of Ti-6Al-4 V alloy.* The International Journal of Advanced Manufacturing Technology, 2016. **87**(1-4): p. 859-866.

219. Caesarendra, W., et al. *Machine degradation prognostic based on RVM and ARMA/GARCH model for bearing fault simulated data*. in *Prognostics and Health Management Conference, 2010. PHM'10.* 2010. IEEE.

220. Paris, P.C., M.P. Gomez, and W.E. Anderson, *A rational analytic theory of fatigue.* The trend in engineering, 1961. **13**(1): p. 9-14.

221. Li, Y., T. Kurfess, and S. Liang, *Stochastic prognostics for rolling element bearings.* Mechanical Systems and Signal Processing, 2000. **14**(5): p. 747-762.

222. Whitle, P., *Hypothesis testing in time series analysis*. Vol. 4. 1951: Almqvist & Wiksells.

223. Haykin, S., *Adaptive filter theory, 1996.* telecommunication systems, radio resource management,(adhoc) multihop relay system, sensor network and particularly their applicable issues to 4G mobile communication systems and cognitive radio systems. Bath in, 2000: p. 12-13.

224. So, C.F., S.C. Ng, and S.H. Leung, *Gradient based variable forgetting factor RLS algorithm.* Signal Processing, 2003. **83**(6): p. 1163-1175.

225. Zhang, S. and J. Zhang. *An RLS algorithm with evolving forgetting factor*. in *Signal Design and its Applications in Communications (IWSDA), 2015 Seventh International Workshop on*. 2015. IEEE.

226. Group, M.R.W., *Report of large motor reliability survey of industrial and commercial installations, Part I.* IEEE Trans. Ind. Appl, 1985. **21**(4): p. 853-864.

227. Kotzalas, M.N. and T.A. Harris, *Fatigue failure progression in ball bearings.* TRANSACTIONS-AMERICAN SOCIETY OF MECHANICAL ENGINEERS JOURNAL OF TRIBOLOGY, 2001. **123**(2): p. 238-242.

228. Tong, W., *Mechanical design of electric motors*. 2014: CRC press.

229. Harris, T. and W.K. Yu, *Lundberg-Palmgren fatigue theory: Considerations of failure stress and stressed volume.* Journal of Tribology, 1999. **121**(1): p. 85-89.

230. Gebraeel, N., et al., *Residual life predictions from vibration-based degradation signals: a neural network approach.* IEEE Transactions on industrial electronics, 2004. **51**(3): p. 694-700.

231. Soualhi, A., et al., *Prognosis of bearing failures using hidden Markov models and the adaptive neuro-fuzzy inference system.* IEEE Transactions on Industrial Electronics, 2014. **61**(6): p. 2864-2874.

232. Liu, L. and J. Wan, *A study of Shanghai fuel oil futures price volatility based on high frequency data: long-range dependence, modeling and forecasting.* Economic Modelling, 2012. **29**(6): p. 2245-2253.

233. Barndorff‐Nielsen, O.E. and N. Shephard, *Econometric analysis of realized volatility and its use in estimating stochastic volatility models.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 2002. **64**(2): p. 253-280.

234. Andersen, T.G., et al., *Modeling and forecasting realized volatility.* Econometrica, 2003. **71**(2): p. 579-625.

235. Corsi, F., *A simple long memory model of realized volatility.* 2004.

236. Shirota, S., T. Hizu, and Y. Omori, *Realized stochastic volatility with leverage and long memory.* Computational Statistics & Data Analysis, 2014. **76**: p. 618-641.

237. Li, R., P. Sopon, and D. He, *Fault features extraction for bearing prognostics.* Journal of Intelligent Manufacturing, 2012. **23**(2): p. 313-321.

238. Chen, Y., B. Yang, and J. Dong, *Time-series prediction using a local linear wavelet neural network.* Neurocomputing, 2006. **69**(4-6): p. 449-465.

239. Pindoriya, N., S. Singh, and S. Singh, *An adaptive wavelet neural network-based energy price forecasting in electricity markets.* IEEE Transactions on Power Systems, 2008. **23**(3): p. 1423-1432.

240. Adamowski, J. and H.F. Chan, *A wavelet neural network conjunction model for groundwater level forecasting.* Journal of Hydrology, 2011. **407**(1-4): p. 28-40.

241. Berenji, H.R. and Y. Wang. *Wavelet neural networks for fault diagnosis and prognosis*. in *Fuzzy Systems, 2006 IEEE International Conference on*. 2006. IEEE.

242. Zhang, Z., Y. Wang, and K. Wang, *Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network.* Journal of Intelligent Manufacturing, 2013. **24**(6): p. 1213-1227.

243. Vachtsevanos, G. and P. Wang. *Fault prognosis using dynamic wavelet neural networks*. in *AUTOTESTCON Proceedings, 2001. IEEE Systems Readiness Technology Conference*. 2001. IEEE.

244. Lu, Y., et al., *Physics-Embedded Machine Learning: Case Study with Electrochemical Micro-Machining.* Machines, 2017. **5**(1): p. 4.

245. Harris, T.A., *Rolling bearing analysis*. 2001: John Wiley and sons.

246. Sawalhi, N., W. Wang, and A. Becker, *Vibration signal processing for spall size estimation in rolling element bearings using autoregressive inverse filtration combined with bearing signal synchronous averaging.* Advances in Mechanical Engineering, 2017. **9**(5): p. 1687814017703007.

247. Eshel, G., *The yule walker equations for the AR coefficients.*

248. Lewis, M. and B. Tomkins, *A fracture mechanics interpretation of rolling bearing fatigue.* Proceedings of the Institution of Mechanical Engineers, Part J: Journal of Engineering Tribology, 2012. **226**(5): p. 389-405.

249. Camci, F., et al., *Feature evaluation for effective bearing prognostics.* Quality and reliability engineering international, 2013. **29**(4): p. 477-486.

250. Doucoure, B., K. Agbossou, and A. Cardenas, *Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data.* Renewable Energy, 2016. **92**: p. 202-211.

251. Okkan, U., *Wavelet neural network model for reservoir inflow prediction.* Scientia Iranica, 2012. **19**(6): p. 1445-1455.

252. Skelton, R., T. Vilhelmsen, and G. Webster, *Energy criteria and cumulative damage during fatigue crack growth.* International Journal of Fatigue, 1998. **20**(9): p. 641-649.

253. McGeough, J.A., *Principles of electrochemical machining*. 1974: CRC Press.

254. Kozak, J., *Thermal models of pulse electrochemical machining.* Bulletin Of The Polish Academy Of Science, Technical Sciences, 2004. **52**(4).

255. Lee, S. and X. Li, *Study of the effect of machining parameters on the machining characteristics in electrical discharge machining of tungsten carbide.* Journal of materials processing Technology, 2001. **115**(3): p. 344-358.

256. Yue, X., et al., *Thermal, mechanical and chemical material removal mechanism of carbon fiber reinforced polymers in electrical discharge machining.* International Journal of Machine Tools and Manufacture, 2018. **133**: p. 4-17.

257. Klocke, F., et al., *The effects of powder suspended dielectrics on the thermal influenced zone by electrodischarge machining with small discharge energies.* Journal of materials processing technology, 2004. **149**(1-3): p. 191-197.

258. Yadav, V., V.K. Jain, and P.M. Dixit, *Thermal stresses due to electrical discharge machining.* International Journal of Machine Tools and Manufacture, 2002. **42**(8): p. 877-888.

259. Bose, A., et al., *A novel, self-expanding, nitinol stent in medically refractory intracranial atherosclerotic stenoses: the Wingspan study.* Stroke, 2007. **38**(5): p. 1531-1537.

260. Pelton, A., et al., *Effects of thermal cycling on microstructure and properties in Nitinol.* Materials Science and Engineering: A, 2012. **532**: p. 130-138.

261. Dubey, A.K. and V. Yadava, *Laser beam machining—A review.* International Journal of Machine Tools and Manufacture, 2008. **48**(6): p. 609-628.

262. Parandoush, P. and A. Hossain, *A review of modeling and simulation of laser beam machining.* International journal of machine tools and manufacture, 2014. **85**: p. 135-145.

263. Yang, J., et al., *Experimental investigation and 3D finite element prediction of the heat affected zone during laser assisted machining of Ti6Al4V alloy.* Journal of Materials Processing Technology, 2010. **210**(15): p. 2215-2222.

264. Lu, Y., et al., *Physics-embedded machine learning: case study with electrochemical micro-machining.* Machines, 2017. **5**(1): p. 4.

265. Lohrengel, M., K. Rataj, and T. Münninghoff, *Electrochemical Machining—mechanisms of anodic dissolution.* Electrochimica Acta, 2016. **201**: p. 348-353.

266. Bhattacharyya, B., J. Munda, and M. Malapati, *Advancement in electrochemical micro-machining.* International Journal of Machine Tools and Manufacture, 2004. **44**(15): p. 1577-1589.

267. Thorpe, J. and R. Zerkle, *Analytic determination of the equilibrium electrode gap in electrochemical machining.* International Journal of Machine Tool Design and Research, 1969. **9**(2): p. 131-144.

268. Zain, A.M., H. Haron, and S. Sharif, *Prediction of surface roughness in the end milling machining using Artificial Neural Network.* Expert Systems with Applications, 2010. **37**(2): p. 1755-1768.

269. Fu, Y., et al., *Analysis of feature extracting ability for cutting state monitoring using deep belief networks.* Procedia Cirp, 2015. **31**: p. 29-34.

270. Li, Z., Y. Wang, and K. Wang, *A data-driven method based on deep belief networks for backlash error prediction in machining centers.* Journal of Intelligent Manufacturing, 2017: p. 1-13.

271. Kim, D.-H., et al., *Smart machining process using machine learning: A review and perspective on machining industry.* International Journal of Precision Engineering and Manufacturing-Green Technology, 2018. **5**(4): p. 555-568.

272. LeCun, Y., et al., *Gradient-based learning applied to document recognition.* Proceedings of the IEEE, 1998. **86**(11): p. 2278-2324.

273. Srivastava, N., et al., *Dropout: a simple way to prevent neural networks from overfitting.* The Journal of Machine Learning Research, 2014. **15**(1): p. 1929-1958.

274. Zou, P., et al. *Electrochemical Micro-Machining Process Parameter Optimization Using a Neural Network-Genetic Algorithm Based Approach*. in *Proceedings of the International Conference on Manufacturing Technologies, San Diego, CA, USA*. 2017.